

行政院國家科學委員會專題研究計畫 成果報告

台灣棒球運動珍貴新聞檔案數位資料館之建置 (II)

計畫類別：個別型計畫

計畫編號：NSC94-2422-H-032-002-

執行期間：94 年 03 月 01 日至 95 年 02 月 28 日

執行單位：淡江大學資訊與圖書館學系

計畫主持人：林信成

計畫參與人員：康珮熏、陳瑩潔、游忠諺

報告類型：完整報告

處理方式：本計畫可公開查詢

中 華 民 國 95 年 3 月 29 日

行政院國家科學委員會補助專題研究計畫成果報告

台灣棒球運動珍貴新聞檔案數位資料館之建置(II)

計畫類別：個別型計畫

計畫編號：NSC 94-2422-H-032-002

執行期間：94 年 3 月 1 日至 95 年 2 月 28 日

計畫主持人：林信成

共同主持人：

計畫參與人員：游忠諺、康珮熏、陳瑩潔

成果報告類型：完整報告

處理方式：本計畫可公開查詢

執行單位：淡江大學資訊與圖書館學系

中 華 民 國 九 十 五 年 三 月 二 十 日

目次

目次.....	1
第一章 導論.....	2
1.1 研究動機與背景.....	2
1.2 研究目的.....	2
1.3 研究方法與步驟.....	2
第二章 數位化流程.....	4
2.1 流程設計與規劃.....	4
2.2 數位化產出.....	7
第三章 相關研究.....	8
3.1 專卷加值.....	8
3.2 第一年度計畫成果.....	9
3.3 Metadata 標準規格探析.....	10
3.4 OAI-PMH	13
3.5 加值系統.....	14
第四章 Metadata 轉換與 OAI-PMH 系統設計實作.....	17
4.1 新聞 Metadata 比較分析.....	17
4.2 新聞 Metadata 轉換系統實作.....	34
4.3 OAI-PMH 檢索系統設計	36
第五章 Wiki 協作系統的應用.....	43
5.1 系統建置與平台技術.....	44
5.2 知識組織與內容加值.....	46
5.3 社群建構與管理策略.....	49
第六章 結論.....	52

第一章 導論

1.1 研究動機與背景

中華棒球代表隊每次進軍國際棒球比賽，皆在舉國關注的氛圍中掀起的一波波棒球熱潮，再度顯示棒球運動在台灣紮根之深、影響之遠，的確已成為台灣文化不可分割的一部份，更是台灣老、中、青不同時代的共同記憶！

雖然棒球已經被公認為台灣的「國球」，然而，長期以來大多數棒球界人士所感嘆的卻是：「台灣是個發展棒球卻不紀錄棒球的地方」，這也成了台灣棒球運動的一大遺憾。為此，近年來已陸續有研究者開始投入研究台灣棒球史的行列，試圖為台灣棒球的輝煌歷史留下記錄，不過，卻鮮少有從資料數位典藏與資訊加值利用的角度出發者。

本計畫乃與聯合知識庫合作，將聯合報自光復以來所珍藏與台灣棒球運動相關之新聞資料，於第一年度已完成新聞專卷加值，包括文字、照片以及眾多經專業加值整理的珍貴資料，進行內容數位化之工作；並完成NRMF格式Metadata轉換與DAC格式Metadata大批匯出至聯合目錄等成果，使其成為極具參考價值之數位典藏庫。

今年第二年度計劃針對Metadata格式轉換及資訊加值與應用等面向做更進一步的深入探討，以提供使用者更多元化的資訊需求。此外，也改善現有之檢索系統，使其功能與介面更具親和力。並著手建立深度加值並免費開放外界使用之「台灣棒球維基館」，以求協助棒球界人士、棒球運動研究者以及所有對棒球有興趣的社會大眾，能更深入、更有效率的探索台灣棒球運動與社會發展的脈動。

1.2 研究目的

本計畫主要與聯合報系合作，對其所珍藏之「台灣棒球運動珍貴新聞專卷檔案」進行圖文資料內容數位化之工作，目的在於以數位典藏方式保存台灣珍貴的棒球史料，並提供加值利用。

主要將探討適合電子新聞資料庫所適用的Metadata及全文標示技術，以促進系統對資料之檢索效能，便於管理與利用。並運用知識組織概念，彙整台灣棒球新聞與其相關資料，針對台灣棒球各類主題延伸出多樣化的加值應用，以彌補原始棒球新聞資料庫提供使用者閱讀時的不足之處。

同時也運用圖書館界的分類與編目理念導入數位化新聞資料庫的可行作法。過去報紙文獻尚未數位化的年代，讀者必須在成堆泛黃的舊報紙中翻找相關資料；而現今電子新聞資料庫大都可依賴全文檢索迅速查找，但在目前電子新聞資料庫普遍不著重分類與編目的情況下，讀者仍須在檢索出過多資料後逐一篩選才能找到所需資訊。

1.3 研究方法與步驟

本計畫採系統實作法，先將各種新聞Metadata格式進行了解，選取具特色四

者進行比對分析，在進行建置整個Metadata轉換系統、OAI-PMH檢索系統及加值服務系統。

在加值服務系統之評估時，以知識組織的概念為出發點，將台灣棒球發展歷程中，萃取出值得記錄的主題為架構，採用Wiki協作系統進行建置，開放讓熱愛棒球的人士能並共同參與編輯，以讓系統資料更趨完善並符合使用者需求。

其整體計畫研究步驟如下：

- (1) 了解各新聞 Metadata 格式，並選取 RSS、DC、NITF 與 DAC 四者進行比對分析各欄位之意義，製作欄位對照表。
- (2) 設計與建置新聞 Metadata 轉換系統。
- (3) 採轉換系統之 Metadata 對應，進行 OAI-PMH 檢索系統之規劃建置。
- (4) 評估加值系統之需求，並分析台灣棒球發展歷程之重要元素，以作為加值系統的資訊架構。
- (5) 建置規劃 Wiki 協作系統，並開放系統，使所有人都可共同參與編輯，並藉此讓系統內容更完整且更符合使用者需求。

本計畫之整體系統如圖 1所示，著重於管理層面與資訊組織。在棒球新聞管理系統中，在原有檢索子系統中加入Metadata轉換模組和顯示模組，並設置RSS訂閱子系統以及OAI-PMH Data Provider檢索系統；後端管理子系統則包含聯合目錄大批匯出模組、新聞過濾整理模組。另外，延伸加值應用建置Wiki加值系統，包括前端子系統包含編輯、顯示、檢索等模組；後端管理子系統，以及RSS訂閱、Metadata轉換等子系統，製作出台灣棒球發展史、台灣棒球大事記、台灣棒球名人錄、台灣棒球一級棒、台灣棒球影像館等各專輯。

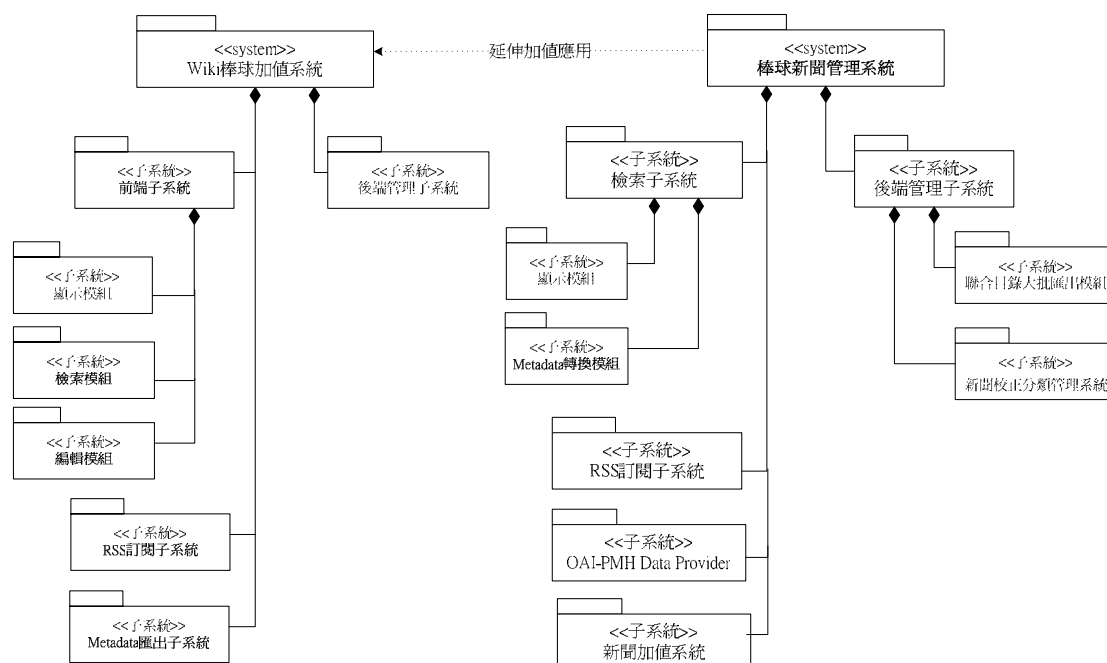


圖 1 本計畫之棒球新聞典藏、管理與加值應用系統整體架構圖（資料來源：本計畫繪製）

第二章 數位化流程

2.1 流程設計與規劃

本計畫之新聞資源為聯合知識庫授權使用，故原始資料數位化乃由聯合知識庫協助完成，共獲取五萬餘篇原始新聞。經由第一年度數位化資料產出的經驗顯示，原始數位化資料的正確性仍有待校正，因此，本年度數位內容著重於提升新聞品質以及簡化分類為目標，以協助使用者獲得更正確的新聞資訊。

本計畫新聞資源之數位化步驟如下：

1. 由聯合知識庫授權提供原始新聞資料；
2. 制訂簡易清楚的單層類目，以滿足使用者資訊需求；
3. 自行建置新聞校正分類管理系統，校正新聞內容並進行模糊分類；
4. 制訂工作流程；
5. 人力資源教育訓練與分配；
6. 數位新聞內容校正與模糊分類實施；
7. 驗收數位新聞校正成果；
8. 將確認無誤的數位化新聞透過 OAI-PMH Data Provider 系統匯入新聞典藏聯合目錄系統

2.1.1 類目規劃

在數位典藏聯合目錄中，規劃以數種不同之「分類架構」進行計畫管理¹。本計畫收錄之台灣棒球體育新聞，原以專卷方式進行新聞事件分類，其原始目的乃為協助報社從業人員查詢資料²，故分類細微，且部分類目以事件為名稱，各類目收錄篇數多寡不一，不但較不適宜一般使用者檢索查詢，對聯合目錄之分類檢索而言也顯得過於繁瑣。

為求類目清楚簡單，我們捨棄起初為求詳盡完整採用主類目之下又再複分的樹狀結構，而改為單層類目，以免新聞類目過於繁多，歸屬不易，不便管理者分類和使用者查找。且資料經歸納整理後，發現棒球新聞事件多以收錄賽事為主，故將各層級賽事獨立成為類目名稱，其餘類目則以較泛稱之描述為名，總共分為十四類。本計畫之分類架構及其對應至數位典藏聯合目錄之整體架構如圖 2 所示。此外，因本計畫以新聞資訊為主，故單篇新聞事件採取多重分類方式，以滿足使用者從不同角度搜尋新聞資訊，在原有系統中增設分類管理功能以便於管理者使用。

¹ 數位典藏國家型科技計畫_內容發展分項計畫：聯合目錄系統建置子計畫，「數位典藏聯合目錄分類架構」，可得自<<http://catalog.ndap.org.tw/dacs4/System/Catalog/Catalog.jsp>>(2005/07/27)。

² 孫正宜，「新聞專卷的數位化與加值應用—以台灣棒球報紙新聞數位典藏為例」，淡江大學資訊與圖書館學系研究所碩士論文，民 93。

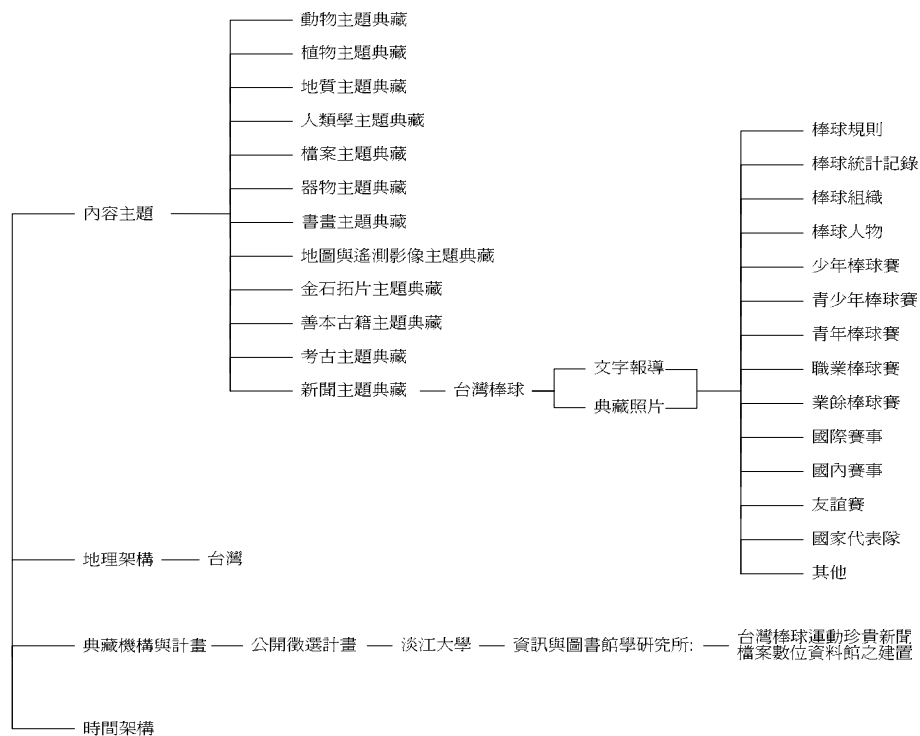


圖 2 本計畫對應於聯合目錄之分類架構 (資料來源：本計畫設計繪製)

2.1.1 後台建置

本計畫所建置之新聞校正分類管理系統，以帳號密碼作人員限定，主要包含新聞瀏覽修改與模糊分類兩大功能。其管理流程為如圖 3 所示，數位化新聞管理人員登入該系統後便可進行新聞內文校正與修改，設定新聞資料處理狀態，確定新聞內容正確無誤之資料依據圖 2 規劃之類目進行模糊分類。

2.1.2 人力資源

本計畫爲了提高棒球新聞資訊的正確率，因此聘請了二十位人員進行數位化新聞校正與分類，並於進行校正工作之前，也針對系統及校正、分類規範舉辦教育訓練說明會，讓修改人員能有對數位新聞進行分類校正有一制性的概念，且在與會中也對於新聞校正分類系統之介面進行討論，以使系統更便於修改人員使用。

2.2 數位化產出

本數位典藏計畫之內容爲聯合報自光復以來長期篩選、整理，與台灣棒球運動相關之報紙新聞專卷檔案，包括文字、照片，以及眾多經過專業加值整理的珍貴資料。

第二年執行至目前爲止之成果如下：

(A) 文字建檔：

本年度持續進行數位新聞文字輸入 15000 則，並結合第一年度 20000 則新聞，共 35000 則進行分類校稿，其細目如下：

- 文字輸入：約 15000 則，約 15.6MB
- 標題建檔：約 15000 則，約 3MB
- 校稿與分類：約 35000 則，約 34.7MB

表 1 每一年度文字新聞之字數與檔案位元數

年代	字數	位元數
1961	24,043	50,514
1962	52,619	108,891
1963	39,551	81,914
1964	21,477	45,531
1965	63,947	130,995
1966	31,184	65,111
1967	41,587	86,935
1968	209,281	429,218
1969	423,427	866,361
1970	494,392	1,013,053
1971	641,628	1,309,135
1972	466,885	954,666
1973	557,117	1,137,243
1974	457,110	938,204
1975	304,251	622,847
1976	387,766	796,427

年代	字數	位元數
1977	343,861	707,491
1978	325,427	667,993
1979	353,755	728,060
1980	273,861	564,313
1981	275,227	567,647
1982	243,644	503,614
1983	212,615	438,674
1984	239,346	493,135
1985	268,685	551,385
1986	278,115	570,854
1987	201,412	416,553
1988	439,338	908,315
1989	316,027	654,999
1990	487,022	1,008,819
1991	588,733	1,217,525
1992	726,567	1,496,412

年代	字數	位元數
1993	629,099	1,293,823
1994	581,485	1,195,660
1995	657,690	1,351,256
1996	1,133,722	2,323,614
1997	893,548	1,831,082
1998	750,063	1,542,343
1999	683,162	1,402,567

年代	字數	位元數
2000	409,402	840,173
2001	613,002	1,257,942
2002	500,882	1,029,011
2003	938,592	1,925,843
2004	104,987	216,202
總和	17,685,534	36,342,350

(B) 照片建檔：

本年度照片掃描 500 張，與第一年度 700 張照片，共計 1200 張進行整理校稿，其細目如下：

- 照片掃描：約 500 張，456MB
- 標題建檔：約 500 則，10KB
- 圖說建檔：約 500 則，約 23,163 字，58KB
- 圖說校稿：約 500 則
- 照片像素調整：約 500 則，97.7MB

第三章 相關研究

在新聞數位化流程啟動的同時，研究小組另外同步進行著 Metadata、分散檢索、資訊深度加值應用等研究課題。本章說明一些相關研究概況和第一年度的研究成果，第四章、第五章則分別說明今年度的研究成果。

3.1 專卷加值

所謂「新聞專卷」是針對不同的新聞主題，把相關的資料加以收集、組織、整理，再分門別類建檔的歷史性新聞文獻。

新聞專卷內容分：過去式、現在式、相關資料。

- 過去式：過去有發生過這類事件嗎？那時是怎麼處理？結果怎樣？
- 現在式：其必要組成要件如下：

五個W，五個W係是人（who）、事（why）、時（when）、地（where）、物（what）。任何新聞事件構成之要素不外以上條件，故專卷或專題也不應脫離這些要素，但是新聞專卷還要包括一個H外加一個P和一個R；

- 一個H係是how：

新聞是有突發事件，但有的新聞在形成前往往會有伏筆，或徵兆。故在做專卷議題時要把為何形成的主要因素列出。

- 一個P係是procedure：

新聞事件的過程是整個專卷議題的主軸，應該是愈詳細愈好，所以在過程的描述要十分的詳細與清楚，這樣的專卷議題日後在供資訊需求者查詢時才有價值。

- 一個R係是result：

新聞專卷來源在報紙，但現在報紙的報導往往因政治因素或版面問題，對一件新聞到最後是不了了之，故整理新聞專卷時要把結果也整理進去。

相關資料：除了網站資料、書目資料，新聞專卷亦包括相關主題的資料。其實所謂相關主題，也是一個獨立主題，只是彼此性質相同，可以相互關聯。

3.2 第一年度計畫成果

本計畫合作單位為聯合報，數位典藏之內容為聯合報自光復以來長期篩選、整理，與台灣棒球運動相關之報紙新聞專卷檔案，包括文字、照片，以及眾多經過專業加值整理的珍貴資料，第一年度(2004)開始執行，預計三年完成，目前第一年已完成之結果如下：

(A) 文字(全年預計約 20,000 篇，每季平均約 5,000 篇)：

- 文字輸入與校稿：約 800 萬字，
- 標題建檔與校稿：全年約 20,000 則（每季平均 5,000 則）

(B) 照片(全年預計約 800 張，平均每季約 200 張)：

- 照片掃描
- 圖說建檔

(C) Metadata 資訊組織：第一年度目前初步採用與 DC 相容的 Metadata 格式(如圖 5 所示)，以及 NRMF。

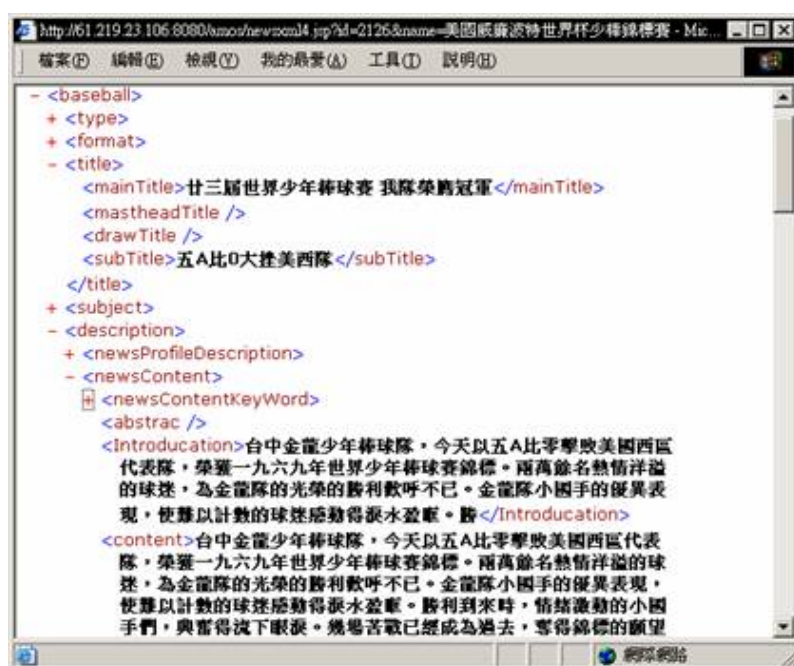


圖 5 第一年計畫採用之 Metadata 格式

(D) 本計畫網站網址為 <http://ndap.dils.tku.edu.tw>。在首頁點選「計畫成果」

即可進入檢索系統，第一年完成初步實驗系統之開發與建置，提供全文檢索與加值檢索兩項功能。如下圖所示。



圖 2 本研究計畫網站首頁

(E) 本計畫第一年度共產出三篇學術論文：

(1) 2004 年 8 月隨中華圖書資訊教育學會至大陸大連理工大學進行學術交流，於「2004 年海峽兩岸圖書資訊學暨教育發展研討會」發表的「報紙新聞專卷的數位典藏與加值利用」；

(2) 擬於 2004 年 11 月淡江大學主辦之「2004 年現代資訊組織與檢索研討會」發表的「數位化新聞的資訊組織」；

(3) 投稿至「數位典藏作業規劃與品質管理研討會」，目前尚在審查中的「提升體育新聞資訊品質之探討－以 SportsML、NITF、NewsML 整合為例」。

此外，本計畫也定期參與內容發展公開徵選計畫辦公室所召開的工作進度執行報告會議，以及新聞主題小組所召開的小組討論會議。

3.3 Metadata 標準規格探析

3.3.1 國際通用 Metadata 格式

傳統的資料編目是以資源為主（Resource-centric），數位資訊的處理則是以關係為主（Relationship-centric），利用Metadata描述紀錄電子資源已成為目前所盛行的方式，因此有許多國際組織積極發展各種類型的Metadata，以方便資訊傳播、典藏。以下將介紹數種國際上新聞資訊常用之Metadata：(1)~(4)為新聞專用；(5)~(6)則為泛用型Metadata，並不限定使用於新聞資訊。

(1) 由國際新聞通訊協會 IPTC（International Press Telecommunication

Council)³所制訂的 NITF (News Industry Text Format)⁴、NewsML (News Markup Language)⁵、SportsML (Sports Markup Language)⁶、ProgramGuideML (Program Guide Markup Language)⁷、EventsML (Events Markup Language)⁸等。IPTC 成立許多專業內容工作小組 (Specialised Content Working Party)，負責旗下相關標準的發展和維護：(a) NITF 著重在新聞內文的描述，主要用於傳統平面媒體的標示，設計初衷在加強新聞檔重覆使用的便利性，讓通訊社、報業、網站等新聞同業可以將新聞稿件轉換成不同的資料格式，並呈現在各種新聞平臺上，達成多重發行的目的。(b) NewsML 的主要作用則是以 Metadata 來封裝多種不同的媒體，主要用於描述電子出版、傳送、典藏的新聞檔，新聞內容的相關描述則藉助 NITF 或其他標示語言嵌入其中；(c) SportsML 主要用於運動項目紀錄的資料交換，以凸顯體育在新聞報導領域的重要地位。(d) ProgramGuideML 是為廣播與電視新聞節目專用的標示語言。(e) EventsML 目前尚處於初步的需求研擬階段，其目的在建立一發佈新聞事件的 XML 格式，以與 NewsML、SportsML 和 NITF 等標準互通。

(2) RSS 是一種用於互通新聞和其他 Web 內容的資料交換規格，RSS 的名稱不甚一致，但都是指同一種 Syndication 的技術，主要可分為三種：Really Simple Syndication(真正簡單的資訊組織格式，版本 RSS 2.0)、RDF Site Summary (基於 RDF 的網站摘錄格式，版本 RSS 0.90 與 RSS 1.0) 與 Rich Site Summary (含意豐富的網站摘錄格式，版本 RSS 0.9x)⁹，是種輕量級的 XML 技術，主要是應用在分享各網站間的新聞標題及其他內容服務的整合¹⁰。使用者可經由「訂閱」(subscribe)的手續，利用 RSS Reader 瀏覽該頻道，獲得即時的新聞資訊，目前已普遍應用於入口引擎、新聞網站、Blog 和 WiKi 等系統中。

(3) PRISM (Publishing Requirements for Industry Standard Metadata) 是一種與出版相關 Metadata 定義與規範標準，其由致力於出版業 XML Metadata 標準之制訂的 IDEAlliance 所發佈。PRISM 包含兩部分一為用於規範 PRISM 架構的規格書，定義了 PRISM 的整體框架，以及 PRISM 聚合者 DTD (PRISM Aggregator DTD)，是應用於多家聯合出版時，用來描述彼此間如何傳送內容及 PRISM Metadata 等 XML DTD 定義文件。

(4) XMLNews¹¹是由 XMLNews.Org 所研擬。XMLNews 可分成兩部份，

³ IPTC, "International Press Telecommunications Council", available from <<http://www.iptc.org/pages/index.php>> (2006/3/5)

⁴ IPTC, "News Industry Text Format", available from <<http://www.nitf.org/>> (2006/3/5).

⁵ IPTC, "News Markup Language", available from <<http://www.newsml.org/>> (2006/3/5).

⁶ IPTC, "Sports Markup Language", available from <<http://www.sportsml.org/>> (2006/3/5).

⁷ IPTC, "Program Guide Markup Language", available from <<http://www.programguideml.org/>> (2006/3/5).

⁸ IPTC, "Events Markup Language", available from <<http://www.iptc.org/EventsML/>> (2006/3/5).

⁹ 林信成、孫正宜，數位化新聞的資訊組織，2004 年現代資訊組織與檢索研討會，頁 69-84，臺北淡江大學，2004/11/19。

¹⁰ 歐坤宗，建構以 RSS 為基礎的新聞內容聯合機制-以《大學報》為例，臺北大學資訊管理研究所碩士論文，民 93，頁 26。

¹¹ XMLNews.Org, "NewsML", available from <<http://xmlnews.org/NewsML/>> (2004/12/20).

XMLNews-Story 用於定義新聞實質內容；XMLNews-Meta 則用於敘述新聞稿件的規格。

(5) 都柏林核心集(Dublin Core, 簡稱 DC)創始於 1995 年,由 OCLC(Online Computer Library Center)和 NCSA(National Center for Supercomputing Application)所一同推動,其非專對數位化新聞而設計的 Metadata,但因其是一個簡易有彈性並且連非專業人員也能輕易使用的 Metadata 格式,所以也可應用於新聞資源描述。

(6) 資源描述框架(Resource Description Framework, 簡稱 RDF)是一種以類似自然語言對資源作描述的架構,作為一整合各種 Metadata 之應用機制,由全球資訊網協會(World Wide Web Consortium, 簡稱 W3C)所制訂。從人工智慧的角度觀之,RDF 相當於「物件-屬性-值」三元素(Object-Attribute-Value Triples)的知識表示法(Knowledge Representation)。目前 RDF 已發展成為建構語意網的核心技術了。

3.3.2 中文新聞 Metadata 格式

中文新聞 Metadata 在近幾年也頗受重視,因此已有多個組織加以研擬新聞專用之 Metadata,以下將介紹數種海峽兩岸新聞資訊常用之 Metadata,(1)~(3)為台灣方面所研發;(4)~(5)則為大陸內地所制定之新聞 Metadata 格式。

(1) 數位典藏聯合目錄 Metadata 格式(Digital Archive Catalog, 本文簡稱為 DAC),乃台灣國科會所制訂,以作為將數位資源匯入聯合目錄之用,適用於中文環境的新聞 Metadata 格式,其主要架構可分為紀錄新聞相關管理資料與新聞事件實質內容兩部份。

(2) 新聞紀錄 Metadata 格式(News Records Metadata Format, 本文簡稱為 NRMF)¹²,乃台灣行政院文建會之國家文化資料庫用新聞 Metadata 格式,適用於中文環境的新聞 Metadata 格式,其主要架構以 DC 的 15 個核心元素為基礎。

(3) 聯合新聞標示語言(UDN Markup Language, 簡稱 UdnML)¹³目前尚在草案階段,僅知以 NewsML 規格為主,大陸新華社的 XinhuaML(後述)為輔,同時也融合了 NITF 可描述新聞內容段落與插題的特色,再配合聯合報系本身之需求制定。

(4) 新華標示語言(Xinhua Markup Language, 簡稱 XinhuaML)¹⁴是大陸新華社所制定,其乃根據新華社內部需求進行研發,以 NewsML 為藍本,參照 NewsML 經驗,針對新華社新聞的採集、加工、發佈和評估等涉及新聞資訊的處理、交換、儲存而制定的一套新聞標示語言。

¹² 行政院文化建設委員會國家文化資料庫知識管理系統, "News Records Metadata Format", 可得自 <<http://km.cca.gov.tw/download/rule.html>> (2005/3/24)。

¹³ 聯合報系於 2004 年 8 月成立 XML 小組,對旗下各報社的新聞格式做規範與 Metadata 的制定。

¹⁴ 新華社技術局標準工組, XinhuaMLv1.0 功能說明書, 2003 年 1 月 18 日。

(5) 中國報業電子新聞文稿格式 (Chinese News Text Format, 簡稱 CNTF)¹⁵, 其為中國報業協會所制訂, 是依據 NITF 所發展, 制定目的在提供中國大陸報業電子新聞文稿基本結構, 以及描述新聞文稿詮釋資料和內容的一套完整的標示語言。

3.4 OAI-PMH

OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting)¹⁶是由 OAI協會 (Open Archives Initiative) 在2001年1 月倡, 乃作為國際上數位化資源的交換標準之一, 在資訊傳播過程中以Metadata提供互通機制, 將分散資源加以匯整, 因此可作為數位圖書館、數位博物館間的通訊協定, 達成分散式數位典藏品整合檢索目的。其主要發展目標為: 1.簡化文件內有效的傳播; 2.提升電子化文件的存取; 3.擴展存取數位化資料種類的範圍。因此, OAI-PMH協定其實作容易、開放性, 採用XML與HTTP等開放標準, 相容性高…等優點, 在歐美已有許多單位進行建置與使用¹⁷。

OAI-PMH主要有五個組成元件¹⁸:

- (1) 資料提供者 (Data Provider): 提供符合 OAI-PMH 協定 Metadata 格式資料;
- (2) 服務提供者 (Service Provider): 經由 OAI-PMH 協定向 Data Provider 取得資料, 並以所獲得的 Metadata 建立加值服務;
- (3) 資料儲存器 (Repository): 透過 HTTP 協定接受 OAI-PMH 命令集, 以回應資料存取需求的資料庫伺服器;
- (4) 資料錄 (Record): 依據 OAI-PMH 協定, 從資料儲存器內將資料以 XML 編碼傳回前端的 Metadata 紀錄。
- (5) 資料集 (Set): 資料儲存器中歸屬於某個資料類別的資料錄所成的集合;

OAI-PMH定義了六個命令集 (Verbs), 其運作模式為Service Provider送出Verbs至Data Provider, 以獲取Repository中的Record。這六個命令集為:

- (1) GetRecord: 檢索 Repository 中單筆的 Metadata 資料錄;
- (2) Identify: 取得 Repository 的識別資訊;
- (3) ListIdentifiers: 取得 Repository 中 Metadata 資料錄識別明細;
- (4) ListMetadataFormats: 檢索 Repository 中所支援的 Metadata 格式;
- (5) ListRecords: 列出 Repository 中指定範圍的所有 Metadata 資料錄;
- (6) ListSets: 檢索 Repository 中的資料集結構。

由於「數位典藏國家型科技計畫」所包含的眾多參與機構與計畫因內容主題

¹⁵ 中國報業協會規範工作組, "中國報業電子新聞文稿格式", 2000 年 5 月。

¹⁶ Open Archives Initiative, "The Open Archives Initiative Protocol for Metadata Harvesting", available from <<http://www.openarchives.org/OAI/openarchivesprotocol.html>> (2005/05/20)

¹⁷ Open Archives Initiative, "OAI community", available from <<http://www.openarchives.org/community/index.html>> (2005/07/24).

¹⁸ 國家圖書館, 「圖書館法專案: 數位圖書館分散式檢索協定」, 可得自<<http://www.lac.org.tw/law/librarylaw/0623/06234.pdf>> (2005/07/24)。

的多樣性，不但分散且相互獨立，為求匯集各計畫之成果，供民眾查詢取用，乃產生以OAI-PMH建置「國家數位典藏聯合目錄」作為服務提供者（Service Provider）之構想，以整合各自分散的眾多數位典藏資訊系統。

3.5 加值系統

為了進行資訊深度加值，第二年決定採用Wiki協作系統作為知識管理平台。因此，以下將針對知識組織、社會性軟體、Wiki協作系統簡介及Wiki協作模式等面向，將其意涵、研究範疇等作初步的整理。至於研究成果則於第五章描述。

3.5.1 社會性軟體（Social Software）

近幾年網路興起開放原始碼型態的社會性軟體的發展，使網路的應用越來越人性化且普及化。此種新一代的網路互動需求，其特性是反應出真實的社會活動與關係的建立。然而，所謂的社會性軟體，是可用來維持個人或團體間的溝通（conversational interaction between individuals or groups）、社會的反饋（social feedback）、人際網絡（social networks）等。¹⁹因此，其是強調運用軟體來作為社交互動的一種模式，如協同創作與溝通等，也發展出幾種常見的社會性軟體，如：IRC、MSN、del.icio.us、Blog、Wiki等。

3.5.2 Wiki 系統簡介

Wiki 系統的創造者Ward Cunningham曾為 Wiki 下了定義，他認為Wiki不只是一套用來儲存與修改資訊的超文件系統，且要每個頁面皆能開放任何使用者編輯，因此必須是一個可自由擴展且相互連結的資料庫系統，只要藉由具備表單功能的瀏覽器即可進行協同創作（Collaboration）。²⁰因此，歸納出其具備下述特點：

- （1）具開放性的協同創作系統：開放所有使用者進行編輯與修改的動作，以合作接力的方式完成某一主題，能使該領域知識越來越完整。
- （2）結構化組織知識：將各個概念或詞彙透過超連結的方式，以網狀形式連結，使知識更趨結構化。
- （3）技術面門檻較低：在系統架設方面，其程式開放原始碼並有多種套件組合可供使用，因此使用者不需花費太多時間於系統建制上，可專心於內容品質的建構，此外，其編輯語法簡易，可使用其特殊語法（如括弧、等號等符號）或搭配 HTML 即可簡單編輯並輸出頁面，亦可在頁面上直接檢視其原始碼即可得知其編輯語法。
- （4）版本控制來維持品質：由於其為開放式編輯，因此其品質或使用者行為較無法掌握，所以記錄下每一筆更動過的記錄，能做有效的品質控

¹⁹ Stowe Boyd, "Are you ready to use social software?" available from <<http://www.darwinmag.com/read/050103/social.html>> (2006/3/5)

²⁰ Bo Leuf, Ward Cunningham, "The Wiki Way: Quick Collaboration on the Web", Addison Wesley, Addison Wesley, 2001.

管。

- (5) 最近更新 (Recent Changes) 的功能有助於編輯與管理：此功能依據更改內容的時間加以排序，並提供近期內更新之內容差異、修改日期、修改幅度、修改者等資訊，並配合特殊標記，以供編輯者與管理者能快速瀏覽並輔助其工作。
- (6) 持續成長的有機體：網頁的內容跟結構因開放編輯的特性，會使其內容越來越多元且豐富。

3.5.2.1 Wiki 協作系統特點

- (1) 開放性協同創作：開放所有使用者對數位內容進行編輯與修改，以合作接力的方式完成某一主題，使該領域知識越來越完善。
- (2) 結構化組織知識：將各個概念或詞彙以類似語意網路之網狀形式連結，並藉由資料庫管理，使知識更趨結構化。
- (3) 技術面門檻較低：在系統架設方面，採開放原始碼作法，並有多種套件組合可供選擇，使用者不需花費太多心力於系統開發上，可專心於內容的建構；此外，其編輯語法簡易，非常容易學習。
- (4) 以版本控制維持品質與安全：由於採開放式編輯，因此系統通常會記錄下每一筆異動，以便發生未預期之變動時，能迅速回復到先前的版本。
- (5) 最近更新有助於編輯與管理：依據內容的更改時間排序，提供近期內更新之差異、修改日期、修改幅度、修改者等資訊，並配合特殊標記，使編輯與管理工作更容易。
- (6) 持續成長的有機體：因開放編輯的特性，理想上，Wiki的內容與結構將隨著時間成長，越來越多元且豐富。

3.5.2.2 Wiki 協作系統缺點

- (1) 開放編輯不易掌控其內容品質：由於其具有人人皆可編輯的特性，內容更改的幅度較大，因此在品質的控管上較難以控制。
- (2) 需大量人力進行管理：無論是內容或社群成員方面的管理，都需要投入大量人力與時間經營。
- (3) 系統功能無法滿足所有需求：沒有一個平台系統是十全十美的，而Wiki系統在檢索與線上人數控管等功能較不足。
- (4) 編碼與安全性問題：Wiki系統是使用UTF-8編碼，其在中文字體的運作會有亂碼問題，此外，其開放的特性也會為其系統安全性帶來另一項危機。

3.5.3 知識組織

知識組織為數位內容加值的方法之一，其理論是建立在知識單元——概念的基礎上，是將無序或分散的特定知識，根據一定的原則與方法，使之有序、集中、

定址，以方便知識的提供、利用和傳播²¹。在網際網路與科技的進步下，最新型態的知識組織為知識處理(Knowledge Processing)的一個過程，主要指知識的分類及構造，其包括：編目(cataloging)、索引(indexing)、分類(clustering and classification)、過濾(filtering)及聯結(linking)²²。

隨著時代與科技的演進，人類知識組織系統也逐漸產生變動，而所謂知識組織的理論是建立在知識單元一概念的基礎上，知識是以知識單元（即概念）及許多語詞或句子的可能組合來表達，換句話說，知識組織是將無序或分散的特定知識，根據一定的原則與方法，使之有序、集中、定址，以方便知識的提供、利用和傳播²³。也就是說其目的是希望使用者能經由知識組織而達到便於檢索、瞭解、利用及共享的機制，而以下將針對知識組織概念的相關概說作一簡介，也可從中體會無論是過去的紙本型式當道，至現今數位型式的崛起，在知識組織的範疇裡，改變的是載體的轉換而其基本概念仍是不變的。

3.5.3.1 分類架構

分類的概念在傳統的圖書分類編目(Cataloging and Classification)的角度來說，是將圖書資料的形體及內容特性，按照一定的編目規範記載下來，以便讀者認識、辨別並利用。然而演進於現代的資訊組織(Information Organization)，是在檢視所有資訊類型及格式之組織原則、概念、理論及方法，主要的資訊組織方式包括：書目學、分類編目、索摘及記錄管理等。²⁴

3.5.3.2 控制詞彙

控制詞彙的理念在圖書館界中運用得十分廣泛，其並非始於線上目錄的運用，在書本及卡片目錄時代即有權威控制的觀念與具體作法。它是指將詞與詞之間依照同義關係、層級關係及其他關係相互編輯起來，其主要的功能是提供一個標準化的詞彙，供資訊儲存與檢索利用。

根據美國國家標準ANSI/NISO Z39.19-1993 (Guidelines for the Construction, Format, and Management of Monolingual Thesauri) 指出：將控制詞彙利用已知的清楚地展現詞彙間同義、同形異義、層級和關聯的關係，而這些關係也需反指使用。其目標是促進文件標引的一致、控制後組合資訊儲存與檢索系統、藉由款目詞與描述語的連結而便於搜尋²⁵。

此外，陳昭珍教授也曾針對控制詞彙提及其優點是（1）可解決很多語意上的問題；（2）可表達主題之類屬（上層主題、下層主題、相關主題）間的關係；

²¹ 阮明淑、溫達茂，「ontology 應用於知識組織之初探，佛教圖書館館訊」，2002 年 12 月，32 期，頁 6。

²² 陳和琴，「資訊組織與編目規範」，可得自<<http://datas.ncl.edu.tw/catweb/datas/3-1-007-1.pdf>> (2005/07/15)

²³ 同註 21。

²⁴ 同註 22。

²⁵ 數位典藏國家型科技計畫後設資料工作組，「建構數位典藏『主題索引典』：簡介與實務」，2002 年 3 月，頁 1。

(3) 是一種知識地圖的呈現，其缺點是(1) 成本高；(2) 所包含的範圍未必適當(3) 會有人為的錯誤(4) 可能有過時的語詞，新詞彙的加入速度太慢；(5) 詞彙間的關係難以系性的建立。

3.5.3.3 索引典

索引典也是控制詞彙的另一種呈現方式，於1947-1950期間Mooer 先後提出描述詞、資訊檢索、資訊檢索系統等術語為索引典的發展奠定思想基礎，並於1957年由H. Brownson 第一次正式使用索引典這個術語²⁶。它是以代表科學概念的關鍵詞為中心，將關鍵詞之同義概念、上位概念、下位概念及聯想概念同時揭示出來的系統。在結構上，索引典是以關鍵詞為中心，因其上位及下位概念可同時展開，不受等級分類系統中，同一層級只能用一個劃分標準的限制。

根據ANSI/NISO Z39.19-1993將其界定有下列四種目的(1) 轉換：將自然語言轉換成控制詞彙；(2) 一致性：促進標引詞的一致性；(3) 指示關係：指出詞彙間的語意關係(4) 檢索：作為檢索文件時的搜尋輔助；因此，索引典可做到下述六項指引²⁷：

1.在特定的知識領域，提供概念之間的關係指引，幫助標引人員（indexer）及檢索人員（searcher）了解該學科的知識架構。使用者並可藉著詞間關係，由已了解的知識概念來了解新概念的涵義。

2.在特定的知識領域，提供標引人員一套標準化詞彙，確保資料標引的一致性。

3.為了確保資料處理的一致性，索引典為一概念一詞，而其他同義詞則以用代關係指引。

4.詞與詞之間有明確的詞間關係存在，這些關係能區分各詞在語義網架構中的功能，可作為使用者選擇正確詞彙的指引。

5.當查出的資料太多或太少時，索引典的層級結構可幫助使用者擴大或縮小檢索主題的詞彙範圍。

6.最好能成為該學科的標準用語

第四章 Metadata 轉換與 OAI-PMH 系統設計

4.1 新聞 Metadata 比較分析

在前述DC、DAC、RSS、PRISM、RDF等眾多的新聞Metadata規格中，經分析評估後，我們選用DC、DAC、RSS和NITF四種Metadata格式作為實作轉換機制，其主要原因為：

²⁶ 美國資訊科學學會台灣分會，「索引典理論與實務」引介，可得自
<<http://commdb.nccu.edu.tw/tour/dibl/thintro.html>> (2006/03/16)

²⁷ 黃惠株，「淺談索引典」，佛教圖書館館訊，1996年3月，5，頁2。

(1) DC簡單易用，不但是目前數位資源描述最普遍的Metadata格式，且為其餘若干格式之基礎，使用上具有豐富彈性，搭配RDF可融合各種Metadata，更可配合OAI-PMH協定進行分散性資源檢索，以利資源分享；

(2) DAC為聯合目錄系統共通標準，乃國家數位典藏計畫指定規格，用以整合數位典藏資源，供公眾查詢利用；

(3) RSS為線上新聞訂閱格式，可讓使用者獲得即時新聞資訊，且資料內容以摘要方式呈現，便於使用者過濾所需新聞，以利進一步閱讀；

(4) NITF具豐富的內文語意標示，用於描述新聞內容，其元素規定嚴謹、完整，可詳述5W1H要素。

其餘未於本年度中實作之Metadata格式，則留待作為未來研究方向。以下將針對此四種Metadata之規格進行深入探析。

4.1.1 DC 規格探析

都柏林核心集 (Dublin Core) 的發展目的是希望建立一套描述網路上電子文件特色的方法，協助資訊檢索，用簡單的Metadata記錄來描述種類繁多的電子資源。DC是一個簡單且具有彈性，各種專業人員皆可輕易瞭解、使用的資料描述格式，所以其只規範多數資源類型皆必須提及的資料特性，目前有15個欄位供資料描述，每個元素都為可選元素，也皆可重複。

將DC與RDF架構搭配使用，是描述數位化檔或網路資源的有效方式之一，利用RDF架構可便利於DC與其他Metadata加以整合，以豐富描述該文件之相關資訊，也可適用於數位化新聞資料，有助於數位典藏利用。

兩者搭配使用之規格架構可參考圖 6，其之根元素為<rdf:RDF>在此可宣告所引用之XML Namespace，且<rdf:Description>中之屬性rdf:about為該篇新聞之URI，其餘15個DC元素，則以<dc: title>形式注錄。

將新聞資料內文對應於DC欄位，除了因為DC為最普遍用於紀錄電子資源之Metadata外，未來更可應用與OAI-PMH機制，透過開放式檢索以助於新聞資料檢索與取用。

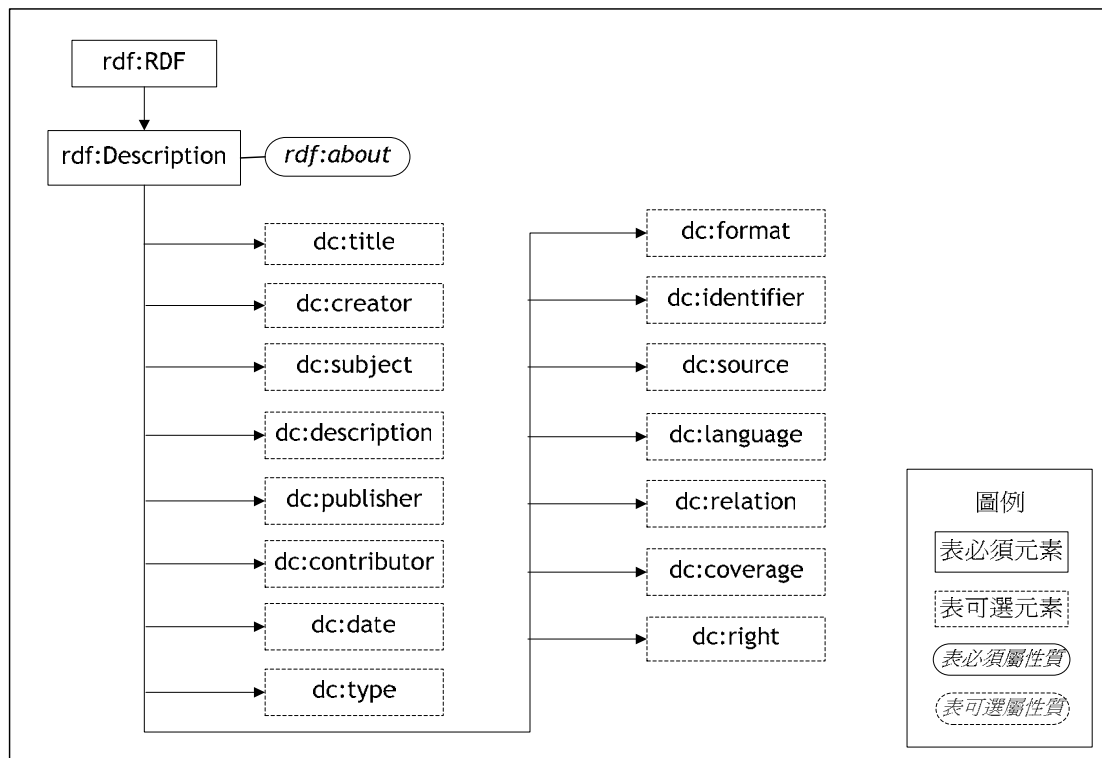


圖 6 RDF+DC 規格樹狀圖 （資料來源：本計畫整理）

4.1.2 DAC 規格探析

數位典藏聯合目錄Metadata格式（Digital Archive Catalog），主要功能乃將數位資源匯入聯合目錄系統之用，其根元素為<DACatalog>，包含<AdminDesc>與<MetaDesc>兩部份：<AdminDesc>為紀錄新聞相關管理資料，該部分由聯合目錄小組自行研訂；<MetaDesc>為紀錄新聞事件實質內容，遵從DC格式，可參閱圖 7。

<AdminDesc>中注錄該研究計畫之匯出機構相關資料，以及關於該計畫之數種分類，如內容架構、地理架構、時間架構等，以作為數位典藏分類檢索之用。其之子元素<DigiArchiveID>為紀錄唯一識別碼，<Hyperlink>則紀錄該資料之link，<ICON>及記載藏品資料的縮圖，以URL表示。

<MetaDesc>則遵循DC的15個欄位，加以注錄該篇新聞之實質內容，如<Title>、<Creator>、<Publisher>、<Language>、<Type>等資料，聯合目錄之DC注錄採取首字大寫，如title改以<Title>為元素名稱。每一個DC tag僅包含一筆描述資料，而非多筆描述資料置入同一DC tag中。

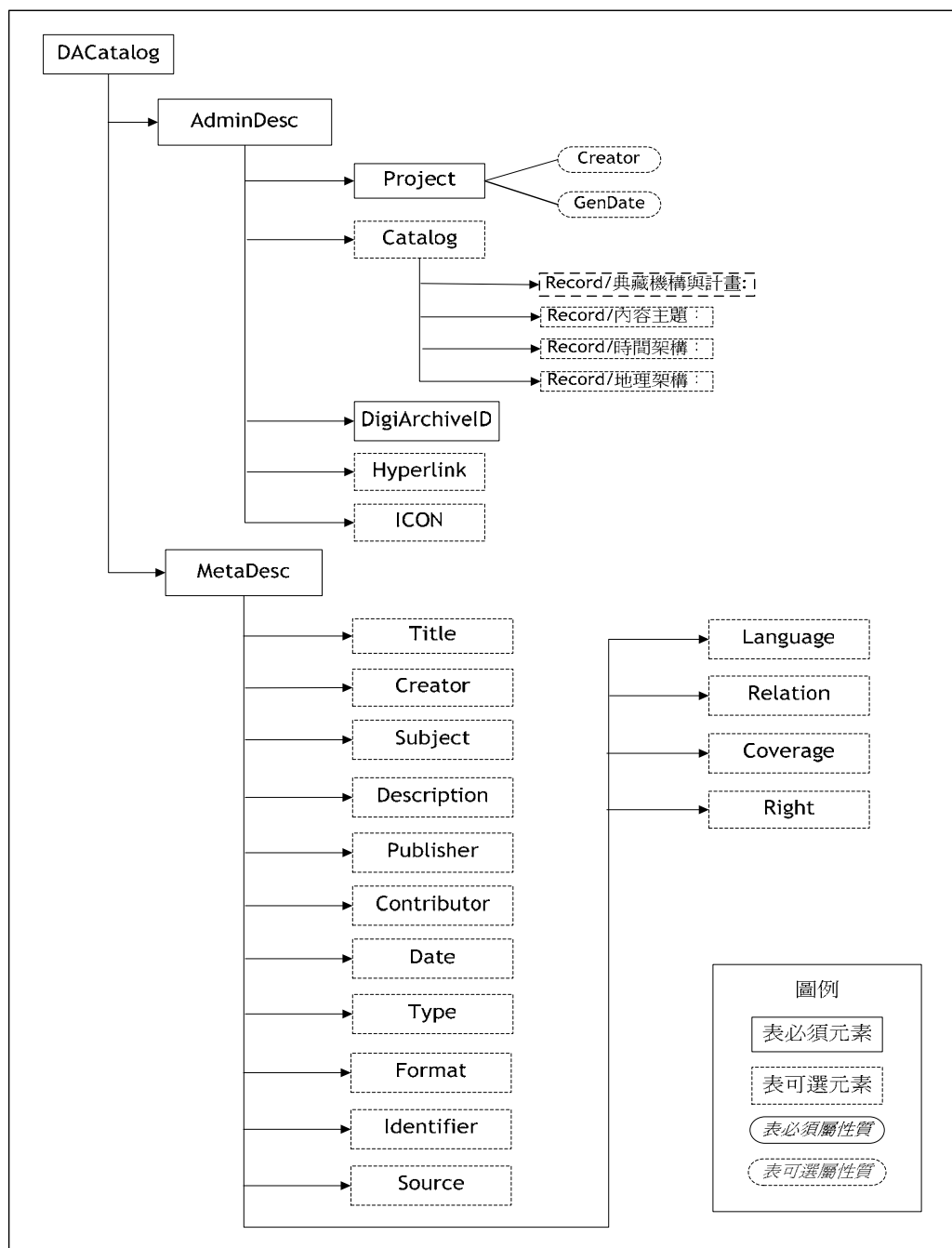


圖 7 數位典藏聯合目錄 Metadata 規格樹狀圖 (資料來源：本計畫整理)

4.1.3 RSS 規格探析

RSS的規格與其發展歷史有很大的關係，從1999年Netscape公司所發展之RSS 0.90版為最早，此版本套用RDF (Resource Description Framework) 框架，根元素為<rdf:RDF>，其規定單一channel只能有15個<item>，各個元素之內容長度也有各別規定，如<channel>之<title>長度只有40字元等規定，使用上較多限制，語法上也較為複雜，因此造成發展群意見產生分歧，令RSS技術衍生出許多不同版本。

當時，部份發展者則認為簡單的規格定義能有助於RSS的推廣與使用，因此UserLand Software在2000年6月衍生發展出RSS 0.91規格²⁸，此版本強調使用上的便利與簡單性，故並無套用RDF框架。RSS 0.91直接以<rss>為根元素，<rss>元素中強制版本屬性(version)用以表示所用的RSS的準確格式，但仍規定單一channel只能有15個item，且<link>與<url>元素規定只能是http或ftp。

2000年12月UserLand公司又發展出RSS 0.92規格²⁹，此版本可視為RSS 0.91之進階版，其規格多相似於RSS 0.91。RSS 0.91與0.92之差別在於新增了五個元素，將新聞資源描述更詳細，並可透過<item>的<enclosure>子元素描述多媒體附件之連結，且不再限制channel中item之數量，以及各元素內容長度之規定³⁰。UserLand公司其後又陸續推出RSS 0.93、RSS 0.94規格，但其修改不大，多依循0.91與0.92規格，故這些UserLand公司所主導之RSS版本可泛稱為RSS 0.9x (Rich Site Summary)，0.9x各版本間皆可相容。

除UserLand公司所發展標榜簡單易用的RSS 0.9x外，非營利網際社群RSS-DEV工作小組也在2000年12月發展出RSS 1.0規格³¹（如圖 8），其語法上繼承Netscape公司發展之RSS 0.90規格，此版本亦套用RDF架構，並使用XML Namespace及延伸性Module語法，此版本之RSS與Netscape 0.90相同涵義，都代表RDF Site Summary。RSS 1.0的根元素為<rdf:RDF>在其中宣告Module Namespace的元素，其RSS 1.0規格中所規定之元素較其他版本少，需搭配Module才可豐富新聞事件之描述。RSS 1.0目前有三種標準Module，即Dublin Core、Syndication與Content，除此尚有16種不同的Propose Modules³²。

²⁸ Dave Winer, "RSS 0.91", available from <<http://backend.userland.com/rss091>> (2005/03/22).

²⁹ Dave Winer, "RSS 0.92", available from <<http://backend.userland.com/rss092>> (2005/03/22).

³⁰ Hammersley, Ben, "Content Syndication with RSS", O'Reilly, 2003, p.18.

³¹ RSS-DEV Working Group, "RDF Site Summary (RSS) 1.0", available from <<http://web.resource.org/rss/1.0/spec>> (2005/1/28).

³² 同註 30, p.72.

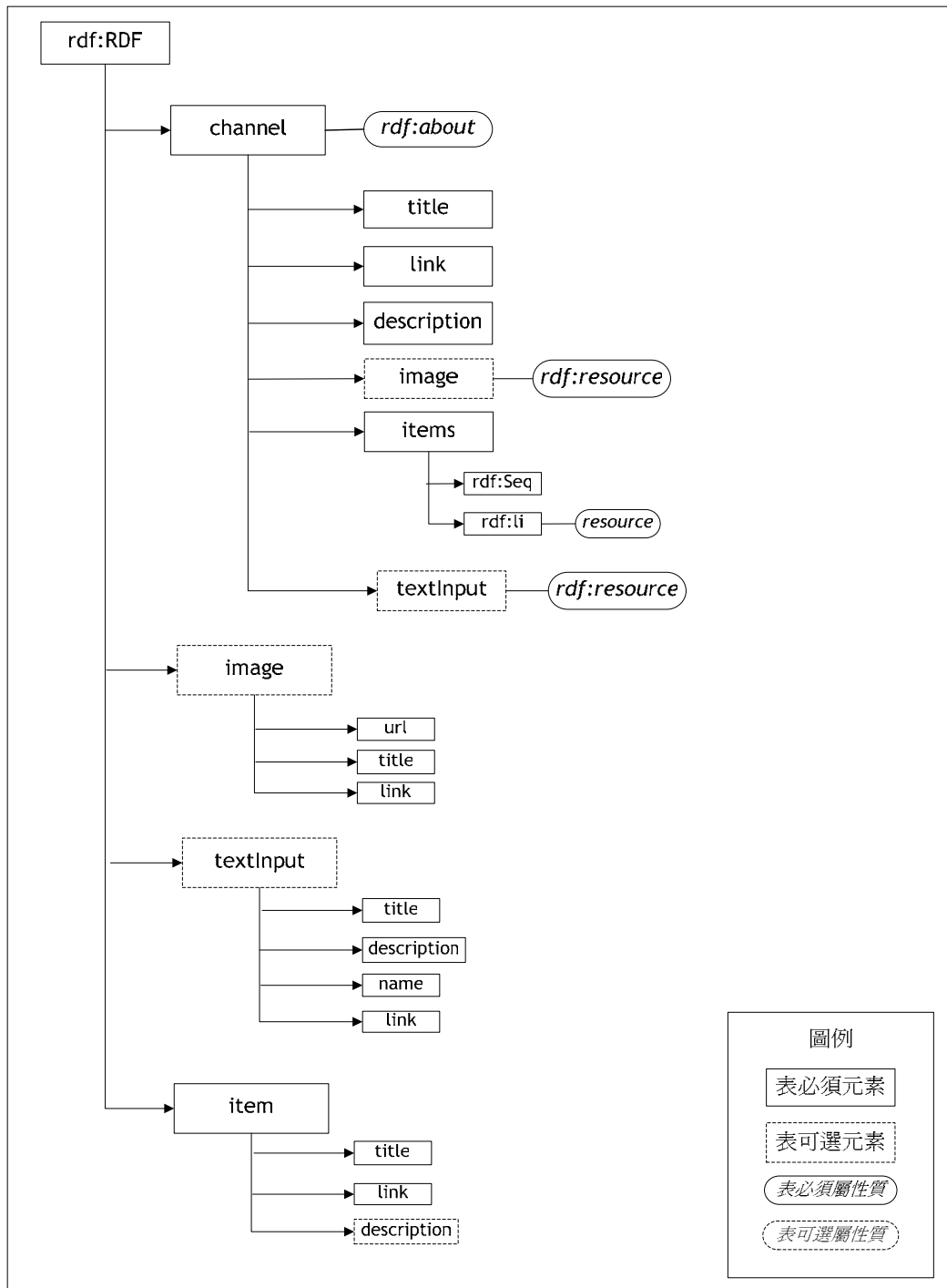


圖 8 RSS 1.0 規格樹狀圖 (資料來源：本計畫整理)

2002年8月UserLand公司再次推出RSS 2.0規格³³ (如圖 9)，此版本針對RSS 1.0的Module延伸架構而產生，基本元素皆繼承RSS 0.9x版本，但加以支援XML Namespace，允許可延伸Module的使用，為與RSS 0.9x版本相互區別，故RSS 2.0之RSS代表Really Simple Summary。其根節點<rss>元素中強制版本屬性(version)改以<rss version="2.0">表示，RSS 2.0以RSS 0.92為基礎，新增許多新的元素，如

³³ Berkman Center, "RSS 2.0 Specification", available from <<http://blogs.law.harvard.edu/tech/rss>> (2004/10/29).

<item>中增加唯一識別子元素<guid>，以解決<url>可能重覆之缺點，<url>不再限定只能為http或ftp，而可支援任何符合URI的格式。RSS 2.0目前已推出一個官方BlogChannel Module，為使RSS Reader可藉以訂閱Blog文章，達到內容聯合。

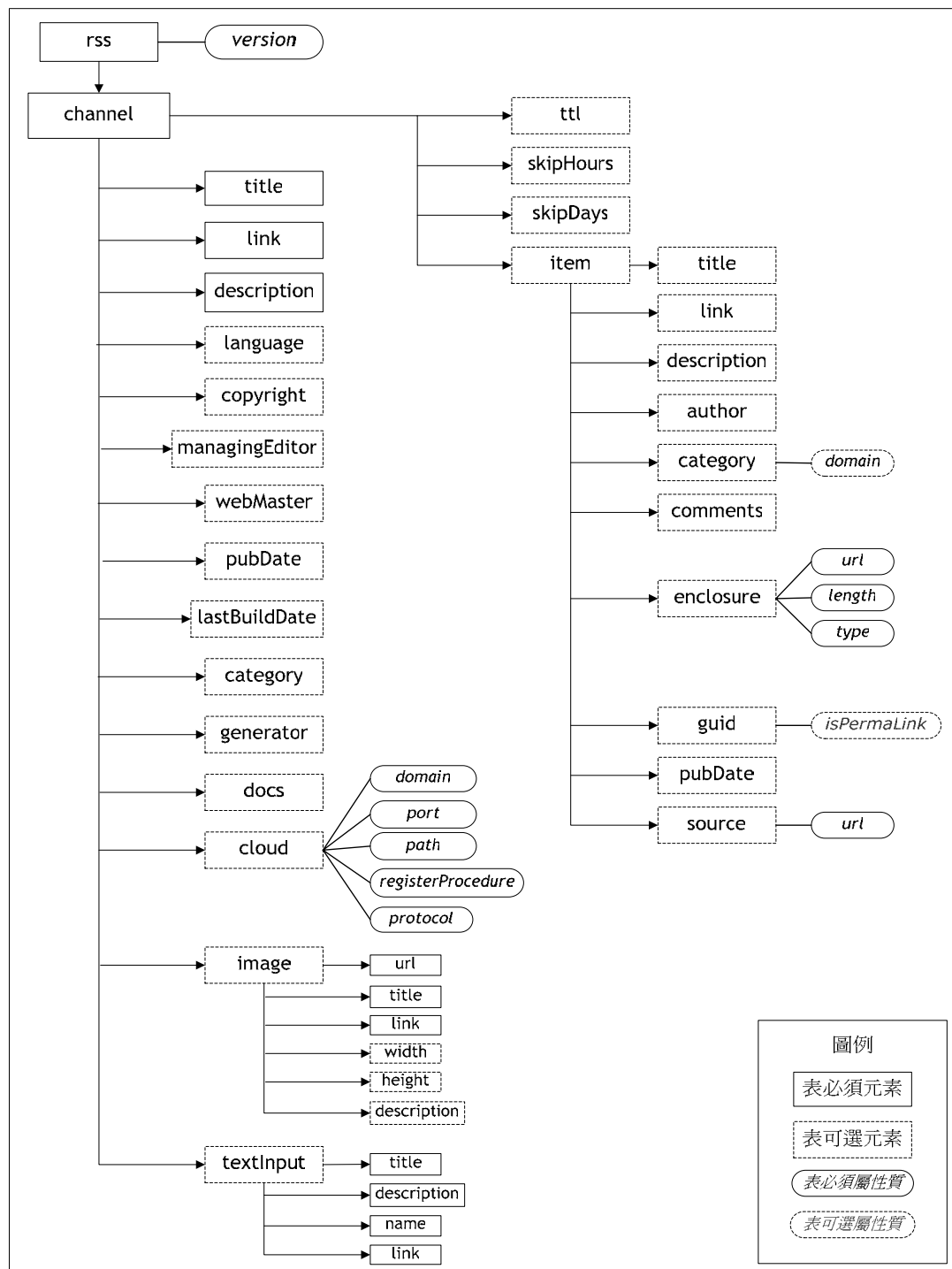


圖 9 RSS 2.0 規格樹狀圖（資料來源：本計畫整理）

RSS 1.0和2.0格式所包含的核心資訊相同，但其結構不一樣，RSS 1.0的根元

素是<rdf:RDF>，RSS 2.0的根元素是<rss>。另一個主要差別是RSS 1.0規格有Namespace限定，RSS 2.0規格就沒有此一限定，但其實兩個版本的格式本質上是一樣的，兩個版本都包含<channel>元素，而<channel>元素又包含三個必須的元素：<title>、<description>和<link>。

目前盛行之RSS版本為具有RDF框架的1.0或簡單使用的2.0版，兩者皆被廣泛利用於Blog、Wiki和網上新聞頻道等，許多知名新聞社網站都提供RSS訂閱支援，而多數RSS Reader也皆可支援RSS 1.0與2.0。

由於Netscape所發展之RSS 0.90版本目前已鮮少人使用，而RSS 0.9x與RSS 2.0一脈相承（如圖 10所示），故本計畫只採取RSS 1.0與RSS 2.0兩規格創建RSS Feed，作為新聞Metadata格式轉換使用。

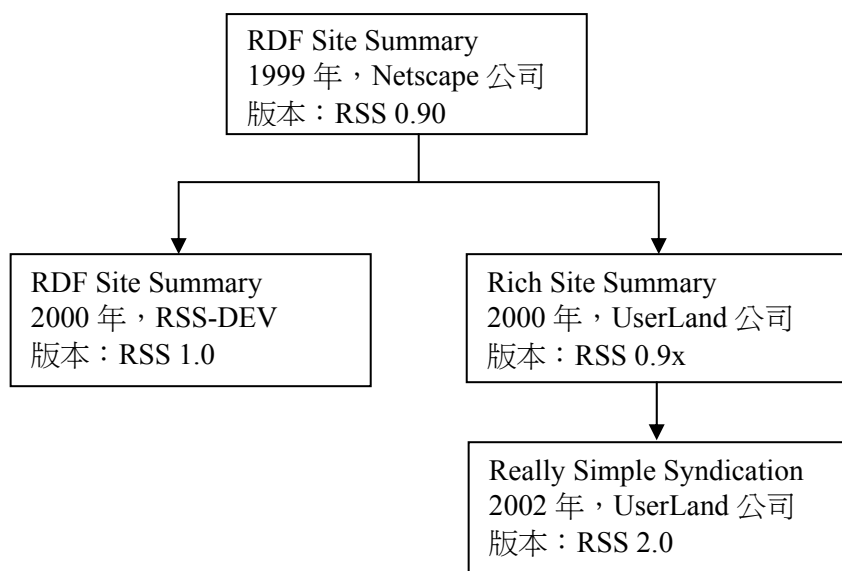


圖 10 RSS 版本發展歷史³⁴（資料來源：本計畫整理）

4.1.4 NITF 規格探析

NITF一開始採用SGML作為其語法規則，直到1998年XML頒佈後，即改採XML語法。NITF以XML為基礎標準，著重於新聞內文的描述，主要用於傳統平面媒體的標示，設計初衷在加強新聞檔重覆使用的便利性，讓通訊社、報業、網站等新聞同業可以將新聞稿件轉換成不同的資料格式，並呈現在各種新聞平臺上，達成多重發行的目的。

NITF規格採用版本為3.2版，包含138個元素，可分為五個層級，規定詳細縝密，將新聞事件完整描述紀錄，其之根元素為<nitf>其下可分為<head>與<body>兩部分，<head>用於紀錄關於該篇新聞相關之資訊，如<title>為該新聞之標題、<toobject>紀錄該新聞之類目、<docdata>可注錄該新聞相關之資料、<pubdata>則以該新聞之發佈相關資訊為內容值等；<body>則紀錄該篇新聞實質之內容資訊，其又可分為<body.head>、<body.content>與<body.end>，<body.head>注錄文章刊

³⁴ RSS 另有 3.0 版本，為純文字檔(text)型式，直接於 IE 瀏覽文字檔或以 mail 流通，該版本屏除 XML、名稱空間(Namespace)以及 HTML 格式，故不在本計畫討論範圍。

頭資訊，如標題<headline>、作者<byline>等；而<body.content>則注錄新聞內容本文資料；<body.end>用於注錄一些與新聞資料相關之訊息，其下有兩個子元素<tagline>與<bibliography>，前者可作為該新聞的最後屬名，該元素之下又可加入許多子元素加強描述性；後者則注錄不限定格式之書目資料。因NITF之元素眾多，層屬關係複雜，因此圖 11只繪出三層元素結構，其所有元素的屬性值與第四、第五層元素架構可參考：

<http://www.nitf.org/IPTC/NITF/3.2/documentation/nitf.html>。

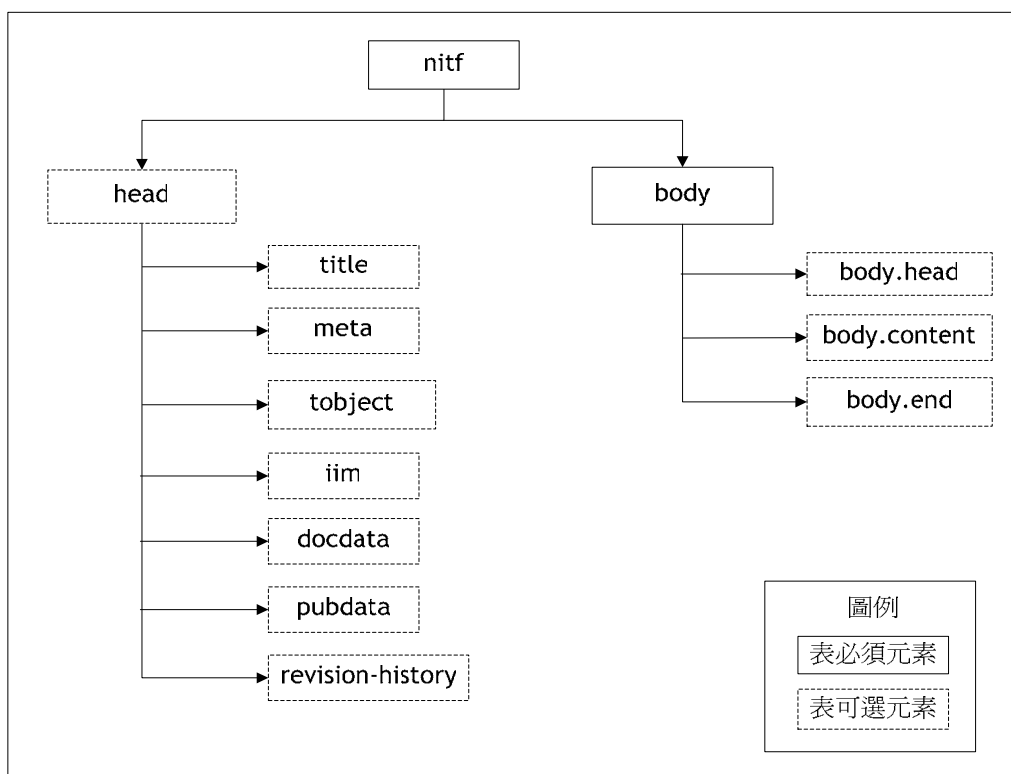


圖 11 NITF 簡易規格樹狀圖 （資料來源：本計畫整理）

4.1.5 四種 Metadata 之綜合比較與對應

表 2為上述四種Metadata之相關資料整理表，由該表可知其中最早發展者為NITF。而四者之個別功能有其相異，如RSS與Dublin Core規格簡單，以描述該新聞資訊之基本資料欄位，使用多彈性，元素與規格較為簡易，且皆可支援延伸性Module的Namespace宣告；NITF為滿足整體新聞媒體描述，使用規範詳細，語法較為複雜；DAC較Dublin Core多了數個作為管理檔使用之元素，但仍繼承Dublin Core之使用彈性與簡易度。四者之共同性在於皆可以text/xml格式，透過通訊協定溝通傳輸，於Web介面呈現。

表 2 四種 Metadata 比較表

項目	Dublin Core	DAC	RSS		NITF
			RSS 1.0	RSS 2.0	
起始	1995 年	2002 年	最初 RSS0.90 為 1999 年		1990 年
發展組織	OCLC 與 NCSA	行政院國家科學委員會	RSS0.90—Netscape		IPTC
			RSS1.0—RSS-DEV	RSS2.0—UserLand	
適用媒體	網路電子資源	數位化電子資源	網路新聞消息		平面新聞媒體
描述重點	定義資源之基本資料	定義資源之管理資料與內容	定義新聞事件之摘要資訊		定義新聞的內容與結構
應用目的	協助典藏描述電子資源	數位典藏資料記錄以供檢索	協助使用者快速便利獲得新聞資訊		協助內容識別
發佈對象	典藏機構	典藏機構	個人用戶		通訊社
規定元素	15 個元素	24 個元素	RSS1.0：23 個元素	RSS2.0：31 個元素	138 個元素
根元素	<rdf:RDF>	<DACatalog>	<rdf:RDF>	<rss>	<nitf>
使用彈性	可搭配 RDF 架構，支援延伸性 Module 的 Namespace 宣告，所有元素皆為可選元素，彈性利用選取所需，皆可重複。	可分為兩部份，管理性資訊需依聯合目錄之規定注錄，數位資源內容描述則沿用 DC 為主要注錄欄位。	RSS1.0—使用 RDF 架構，規定元素較少，支援延伸性 Module 的 Namespace 宣告。	RSS2.0—富彈性，除三個必須元素外其餘皆為可選元素，支援延伸性 Module 的 Namespace 宣告，但並不支援 RDF 語法。	DTD 規定詳盡，階層關係眾多，為求詳細表現新聞資料之內容描述，以滿足新聞加值，不支援延伸性 Module 的 Namespace 宣告與 RDF 語法。
難易	簡單易用	簡單易用	語法較複雜	簡單易用	元素層級架構繁雜
發佈方式	資料交換用，以 XML 於 Web 介面呈現。	資料典藏以作為檢索查詢之用。	拉式（Pull-based）為基礎，使用者主動訂閱 RSS Feed，利用 RSS Reader 瀏覽閱讀。		多做資料交換用，以 XML 於 Web 介面呈現。
格式	皆可透過 text/xml 呈現				
環境	皆建立於網路作業環境，透過通訊協定溝通傳輸				

(資料來源：本計畫整理)

為建置Metadata自動轉換之系統，我們以DC為基礎，與RSS、NITF與DAC三種Metadata之元素，以及資料庫內部Metadata欄位進行比對分析歸納出常用元素與欄位對應關係，作為Metadata轉換之用。

(1) Identifier（資源識別代號）

此元素乃DC用以紀錄該資源之唯一識別碼，故與內部Metadata之主索引相對應，而DAC記述實質內容元素乃以DC為架構，因此該元素與MetaDesc的子元素<Identifier>對應，除此DAC自行規定用以紀錄新聞相關管理元素中也設置<DigiArchiveID>為注錄該檔之唯一識別碼。

在RSS 2.0中<item>的子元素<guid>也是用以表示該item唯一識別字串，但其之格式並不限定使用OSF GUID（Globally Unique Identifier）演算法，故本計畫參考現行之網路新聞發佈方式，給予該元素內容值為該item的URI。而在RSS 1.0中本身規格並無規定唯一識別碼相關的元素，因此需搭配RDF引用DC標準Module，以<dc: identifier>相對應。唯一識別碼屬於關於該篇新聞之相關資訊，而非新聞內容資訊，故在NITF中對應至<head>中紀錄該新聞資料的<docdata>子元素<doc-id>。

(2) Date（出版日期）

Date元素在於DC中並無規定其為該應為該新聞之發佈日期或該XML資源創建的日期，但因本計畫著重於新聞本文之描述，所以將此元素對應於該新聞事件的發佈之日期，故與內部Metadata之新聞日期相對應。DAC格式遵循DC架構，所以相同對應至<MetaDesc>的子元素<Date>，除此DAC自行規定用以紀錄新聞相關管理元素中為作為分類紀錄之用，設置<Catalog>以其之子元素<Record>紀錄不同類別之分類，其中包含「時間架構」之項目，也可相對應之。

RSS 2.0的<channel>子元素<pubDate>是用以紀錄該RSS Feed發布之日期，而<item>的子元素<pubData>是用以表示該item之發佈日期，因此在本計畫中DC的Data應與<item>的子元素<pubData>相對應。RSS 1.0之規格中無規定日期相關元素，但因DC為其標準Module之一，故可搭配RDF與DC本身相同對應至<dc:date>。在NITF中head中紀錄該新聞資料的<docdata>子元素<data.issue>是用以紀錄該新聞發佈的日期；紀錄出版資料的<pubdata>子元素<date.publication>用以紀錄新聞出版日期；而新聞發佈日期也屬於新聞內容的一部份，所以在<body>的<body.head>中設置紀錄該篇新聞之發佈日之元素，於<hedline>的子元素<dateline>，此三者在本計畫中皆可對應於DC的<date>，但其功用不盡相同。

(3) Title（題名）

<title>元素對應至新聞事件，可視為新聞標題，但DC並無區別主標題或副標題於不同元素，雖DC欄位可重複但無法區分主標題與副標題之差異，副標題對於新聞事件又有相當重要性，故本計畫參考現行網路新聞發佈規格，用單一<title>對應於內部Metadata之新聞主標題與副標題，注錄格式為「主標題：副標題」。故DAC也與<MetaDesc>的子元素<Title>相對應，採用相同注錄格式。

在RSS 2.0中記述新聞文件標題者為<item>的子元素<title>，而<channel>的

<title>是用以紀錄該發佈RSS Feed頻道的標題，所以DC之<title>應和<item>的子元素<title>對應，採相同注錄格式。RSS 1.0在<item>元素亦有定義紀錄文章標題用之<title>元素，所以不需引用Module。而NITF為新聞專用內容描述的Metadata，故其格式規定詳細，有不同的元素對應於主標題與副標題，<head>的子元素<title>以及<body>的<body.head>下層<headline>中區分子元素<hl1>用以紀錄該篇新聞之主標題，<hl2>用以紀錄該篇新聞之副標題。

(4) Creator（著者）

DC的該元素是用以紀錄該資源的著者，因此在新聞事件中可對應於撰寫或採訪該新聞之記者，故與內部Metadata之新聞記者對應。DAC記述內文部份以DC為架構，所以DC的<creator>與<MetaDesc>的子元素<Creator>對應。

在RSS 2.0中<author>為紀錄新聞作者的mail，但因本資料庫之內容是報紙新聞，因此採注錄記者姓名而非mail。RSS 1.0本身架構並無規定相關元素，因此需引用DC Module，對應於<dc:creator>。新聞作者也為新聞內容的一部份，所以在<body>刊頭資訊<body.head>之下層<headline>子元素<byline>用以表示該新聞之作者。

(5) Publisher（出版者）

出版者即表示該資源的發行組織，因此DC的<publisher>應注錄發佈該新聞之報別，與內部Metadata之新聞報別相對應。因此在以DC為主架構的DAC中該元素應與<MetaDesc>的子元素<Publisher>對應。

RSS 2.0中並無紀錄出版者之相關元素，但新聞事件之出處應為重要之版權資訊，故本計畫參考現行網路新聞發佈規格，將報別對應至title元素，其注錄方式為「主標題：副標題（報別）」。RSS 1.0本身亦無相關版權資訊注錄元素，因此需引用DC Module對應注錄於<dc:publisher>。新聞報別應為擁有該篇新聞發佈之權利者，也為新聞內容的一部份，可讓讀者得知此一新聞的出處，所以可對應於NITF的<body>刊頭資訊<body.head>中的<headline>子元素<rights>。

(6) Contributor（其他參與者）

DC中該元素用以紀錄除了著者外，對該資源創作有貢獻的其他相關人士或組織，在新聞資料中可對應至攝影者或繪圖者，但因本資料庫的記者欄位將文字記者以及攝影記者等相關作者，紀錄於同一欄位，故本元素與內部Metadata並無對應。但DC的<contributor>可與DAC中的<MetaDesc>的子元素<Contributor>對應。

在RSS 2.0中只有規定<author>一個元素，並無提供參與創作者之相關元素可與之對應。RSS 1.0可引用DC Module加以對應於<dc: contributor>。NITF規格中在<body>刊頭資訊<body.head>的子元素<contributor>用以紀錄資訊的貢獻者，與DC此一元素相對應之。

(7) Description（敘述）

該元素在DC中用以紀錄該文件的摘要或影像資源的內容敘述，對於新聞事件而言，即指新聞之內文，可與內部Metadata之新聞內容相對應。因此在DAC中

也與<MetaDesc>的子元素<Description>對應。

在RSS 1.0與RSS 2.0中記述新聞文件摘要或內容者皆為<item>的子元素<description>，而<channel>的<description>是用以紀錄該發佈RSS Feed頻道的相關內容資訊，故DC之<description>應與<item>的子元素<description>相對應，但在RSS 1.0此元素為可選元素，RSS 2.0此元素則為必備元素，兩者規格有所差別。而NITF紀錄新聞內容者為<body>的子元素<body.content>。

(8) Subject (主題和關鍵詞)

藉由主題與關鍵詞可表示該資源的類目，所以用於報紙新聞資料，可從版名去分別出新聞之主題，因傳統報紙不同版面收集不同主題類目的新聞事件，因此可與內部Metadata之新聞版名對應，若該篇新聞無版名資訊，則由系統注入「體育新聞」作為內容值。DAC格式乃以DC為架構，因此與<MetaDesc>的子元素<Subject>對應。

在RSS 2.0中<item>的子元素<category>是用以注錄該篇新聞的類目；<channel>亦有相同子元素<category>是作為紀錄該頻道的分類，其兩者用法相同，但紀錄內容意義不同，故本計畫的DC之<subject>應與<item>的子元素<category>相對應。RSS 1.0規格無類目相關的元素，需引用DC Module以強化其描述性，所以對應至<dc:subject>元素。NITF有其規定之類目主題層次，NITF的類目紀錄於<head>的子元素<tbody>，有其特別詳細之規定，因此在本計畫NITF並無欄位對應於DC之<subject>。

(9) Source (來源)

在DC中該元素用以紀錄該資源的衍生來源，例如同一作品的不同媒體版本，或者是翻譯作品的來源等，因此本計畫在該元素注錄該篇新聞之HTML格式資料，也就是該篇新聞之URI。在DAC中則以聯合目錄小組自行規定用以紀錄新聞相關管理元素<AdminDesc>中已設置<Hyperlink>作為注錄該新聞之URI，因遵循其典藏系統之規定，故不於<MetaDesc>的子元素<Source>中注錄相同值。

在RSS 1.0與RSS 2.0中紀錄該篇新聞之URI於<item>的子元素<link>，而<channel>的子元素<link>是用以紀錄該頻道之連結，故DC之<source>應與<item>的子元素<link>對應。NITF並非用於網路新聞發佈，因此其之規格無提供相關的注錄元素。

(10) Language (語言)

語言在DC中表示該資源所使用的語言，因此本計畫在該元素注錄該篇新聞所用之語言，可遵循RFC1766的規定來注錄，本計畫之新聞資源皆為繁體中文，故其之固定值為「zh-tw」。而DAC紀錄新聞實質內容元素以DC為主要架構，但因其系統典藏之需求，而改固定值為「中文」，與<MetaDesc>的子元素<Language>對應。

在RSS 2.0中語文相關元素設置於<channel>之下層元素，用以紀錄該頻道的語言，因RSS頻道所用之語言多與<item>所使用的語言相同，所以將DC的<language>與<channel>的子元素<language>對應。RSS 1.0並無語文之相關規定元

素，因此可引用DC Module相對應於<dc:language>。該新聞使用之語言視為新聞內容的一部份，所以NITF之語文相關元素設置於<body>下層之<body.end>的<tagline>其可加入子元素<lang>加以對應，作為該新聞尾端補充資訊。

(11) Rights (版權規範)

<rights>在DC中可表示該資源版權聲明和使用規範，新聞資源而言，應在該元素注錄該篇新聞之版權擁有者，研究之新聞資料皆來自於聯合報系，不同於新聞報別，乃因聯合報系旗下擁有許多不同報別，各個報別為該新聞之出版者，而所有報別皆同歸於聯合報系所擁有其之版權，因此，此元素我們給予固定值為「聯合報系」。DAC則注錄於<MetaDesc>的子元素<Rights>中。

在RSS 2.0中版權相關元素設置於<channel>之下，用以紀錄該頻道的版權資訊，而在本計畫中發布頻道的版權內容即為所有新聞<item>之版權資訊，故該元素與<channel>的子元素<copyright>相對應。RSS 1.0則因有DC為其標準Module之一，所以直接對應於<dc:rights>。如前所述，版權資訊屬於該篇新聞資源的版權擁有者，並非屬於新聞內容資料，故在NITF中對應至head中紀錄該新聞資料的<docdata>子元素<doc.copyright>。

(12) Format (資料格式)

該元素用以紀錄該資源的實體形式或者是數位特徵，此元素多以MIME格式表示，因此在本計畫中將該元素視為注錄該篇新聞之資料格式，而新聞Metadata多以XML格式呈現，故此注錄之固定值為「text/xml」。DAC依DC架構而注錄於<MetaDesc>的子元素<Format>，但依其系統之規定，將固定值改為「XML檔」。

因DC與DAC皆為泛用型的Metadata，各種資源注錄資料格式會有所不同，所以將有規定元素注錄其資料格式，而RSS 2.0、RSS 1.0與NITF格式，皆用於紀錄數位新聞事件，都以text/xml格式呈現，因此三種Metadata本身皆無規定用於紀錄其資料格式之元素。

(13) Type (資源類型)

在DC中該元素用以紀錄該資源的類型或所屬的抽象範疇，因此本計畫在該元素注錄該篇新聞之資源類型，故此在此元素注錄固定值「text」。DAC格式同DC架構，因此<MetaDesc>的子元素<Format>可與之對應，但因聯合目錄系統規定，改其固定值為「文字」。

因DC與DAC皆為泛用型的Metadata，可通用於注錄各種類型資源，所以需將資源類型加以顯示，但RSS 2.0、RSS 1.0與NITF格式，皆為數位新聞事件Metadata專用格式，其資源格式都是text格式，所以三者本身皆無規定相關元素用以紀錄資源類型。

(14) Coverage (涵蓋時空)

DC中以<coverage>作以紀錄關於該資源所涵蓋的時期和地理區域，新聞資源而言，應在該元素注錄該篇新聞之涵蓋時空，而本計畫之新聞資料皆為台灣棒球新聞事件，其時間範圍已在<date>中注錄，故該元素只紀錄涵蓋空間，設定固定值為「Taiwan, ROC」。在DAC新聞實質內容元素<MetaDesc>中以子元素

<Coverage>對應，除此聯合目錄小組自行規定用以紀錄新聞相關管理元素<AdminDesc>中設置<Catalog>作為分類之用，以之下層子元素<Record>用於紀錄不同類別之分類，其中包含「地理架構」之項目，而依其系統之規定，改固定值為「台灣」與之對應。

RSS 1.0與RSS 2.0皆以發布網路新聞為主，故對於時空範圍並無告知的需求，因此其之規格並無規定時空範圍的相關元素。涵蓋時空屬於該篇新聞的新聞資料，也可視為新聞內容的部分，故在NITF在<head>中紀錄該新聞資料的<docdata>子元素<doc-scope>告知該篇新聞之涵蓋範圍，並且在紀錄新聞實質內容<body>的尾端資訊<body.end>下層元素<tagline>中選擇加入子元素<location>以對應。

(15) Relation (關連)

在DC中該元素用以紀錄該資源與其他作品的關連，或所屬的系列和檔案庫，而本計畫將其對應該新聞資源發佈的所屬資料庫，因此其固定值為「<http://ndap.dils.tku.edu.tw>」或「台灣棒球運動珍貴新聞檔案數位資料館」。而DAC除管理元素外，乃以DC為架構，因此該元素與<MetaDesc>的子元素<Relation>對應。

在RSS 2.0中<item>的子元素<source>用以表示該新聞之出處，也就是其發佈之來源資料庫，所以本元素相對應。RSS 1.0則規定DC為其標準Module之一，故與DC本身相同對應至<dc: relation>。新聞所屬之發佈資料庫，應為關於該篇新聞之相關資訊，而非新聞內容資訊，故在NITF中對應至<head>中紀錄該新聞資料的<docdata>子元素<doc.right>。

為建置Metadata自動轉換之系統，我們以DC為基礎，與RSS、NITF與DAC三種Metadata之元素，以及台灣棒球新聞資料庫內部Metadata欄位進行比對分析，歸納出常用元素與欄位對應關係，作為Metadata轉換之用，表 3為比較分析之後的對映結果。

表 3 四種 Metadata 與資料庫欄位對應分析 （資料來源：本計畫比較分析。▶：表示下一層元素；⇒表示該元素屬性值）

外部Metadata					內部Metadata
DC	DAC	RSS 1.0	RSS 2.0	NITF	資料庫欄位
identifier	DACatalog ▶ AdminDesc ▶ DigiArchiveID	rdf:RDF ▶ item ▶ dc:identifier	rss ▶ channel ▶ item ▶ guid	nitf ▶ head ▶ docdata ▶ doc-id⇒id-string	autoid (主鍵)
	DACatalog ▶ MetaDesc ▶ Identifier				
date	DACatalog ▶ MetaDesc ▶ Date	rdf:RDF ▶ item ▶ dc:date	rss ▶ channel ▶ item ▶ pubDate	nitf ▶ head ▶ docdata ▶ date.issue⇒norm	date
				nitf ▶ head ▶ pubdata⇒date.publication	
	DACatalog ▶ AdminDesc ▶ Catalog ▶ Record／時間架構：			nitf ▶ body ▶ body.head ▶ headline ▶ dateline	
title	DACatalog ▶ MetaDesc ▶ Title	rdf:RDF ▶ item ▶ title	rss ▶ channel ▶ item ▶ title	nitf ▶ head ▶ title	headline
				nitf ▶ body ▶ body.head ▶ headline ▶ hl1	
				nitf ▶ body ▶ body.head ▶ headline ▶ hl2	subheadline
creator	DACatalog ▶ MetaDesc ▶ Creator	rdf:RDF ▶ item ▶ dc:creator	rss ▶ channel ▶ item ▶ author	nitf ▶ body ▶ body.head ▶ headline ▶ byline	author
publisher	DACatalog ▶ MetaDesc ▶ Publisher	rdf:RDF ▶ item ▶ dc:publisher	rss ▶ channel ▶ item ▶ title	nitf ▶ body ▶ body.head ▶ headline ▶ rights	paperid
contributor	DACatalog ▶ MetaDesc ▶ Contributor	rdf:RDF ▶ item ▶ dc:contributor	—	nitf ▶ body ▶ body.head ▶ distributor	—
description	DACatalog ▶ MetaDesc ▶ Description	rdf:RDF ▶ item ▶ description	rss ▶ channel ▶ item ▶ description	nitf ▶ body ▶ body.content	document

外部Metadata					内部Metadata
DC	DAC	RSS 1.0	RSS 2.0	NITF	資料庫欄位
subject	DACatalog ▶ MetaDesc ▶ Subject	rdf:RDF ▶ item ▶ dc:subject	rss ▶ channel ▶ item ▶ category	—	pagename
source	DACatalog ▶ AdminDesc ▶ Hyperlink	rdf:RDF ▶ item ▶ link	rss ▶ channel ▶ item ▶ link	—	—
language	DACatalog ▶ MetaDesc ▶ Language	rdf:RDF ▶ channel ▶ dc:language	rss ▶ channel ▶ language	nitf ▶ body ▶ body.end ▶ tagline ▶ lang	—
rights	DACatalog ▶ MetaDesc ▶ Rights	rdf:RDF ▶ channel ▶ dc:rights	rss ▶ channel ▶ copyright	nitf ▶ head ▶ docdata ▶ do c.copyrights⇒ holder	—
format	DACatalog ▶ MetaDesc ▶ Format	—	—	—	—
type	DACatalog ▶ MetaDesc ▶ Type	—	—	—	—
coverage	DACatalog ▶ MetaDesc ▶ Coverage	—	—	nitf ▶ head ▶ docdata ▶ doc-scope⇒scope	—
	DACatalog ▶ AdminDesc ▶ Catalog ▶ Record／空間架構：			nitf ▶ body ▶ body.end ▶ tagline ▶ location	
relation	DACatalog ▶ MetaDesc ▶ Relation	rdf:RDF ▶ item ▶ dc:relation	rss ▶ channel ▶ item ▶ source	nitf ▶ head ▶ docdata ▶ doc.rights⇒owner	—

4.2 新聞 Metadata 轉換系統實作

本系統乃利用JSP與ASP程式撰寫完成，圖 12為檢索子系統與RSS訂閱子系統之活動圖，使用者可經由Web介面選取欲進行之動作，主要可分為：全文檢索、圖片檢索、訂閱歷史上的今天，以及結束作業。

此系統建置目的在於提供一般使用者檢索新聞資源，目前本系統提供文字新聞及圖片新聞兩種資源檢索利用，未來將提供棒球名人錄以及棒球大事記之檢索利用。如圖 12活動圖所示，當使用者選擇全文檢索時，可在查詢介面輸入查詢條件，JSP程式將解讀使用者輸入之條件組合成SQL查詢語法對後端資料庫進行查詢動作，顯示符合條件的多筆文字資料，使用者在進行挑選所需資料全文顯示，再進一步選擇欲顯示之Metadata格式，ASP程式將該篇新聞資訊以被選取之Metadata格式注錄顯示與Web介面上。故該系統之下又分為顯示模組，與Metadata轉換模組。

(1) 顯示模組：本模組是將使用者檢索結果加以顯示給予使用者瀏覽、閱讀，此部份之系統部份於第一年度已建置完成，目前已可供使用者顯示所需之新聞資源。

(2) Metadata 轉換模組：本模組以表 3 所對應之欄位進行 Metadata 轉換，其下包含四種轉換子模組：DC 轉換子模組、DAC 轉換子模組、RSS 1.0/2.0 轉換子模組，以及 NITF 轉換子模組，詳細紀錄新聞資訊，讓使用者可利用此模組選擇欲顯示之 Metadata 格式，透過後端程式運作進行各子模組之轉換，滿足各種不同領域之使用者。各模組之轉換結果可參閱圖 13～圖 17。

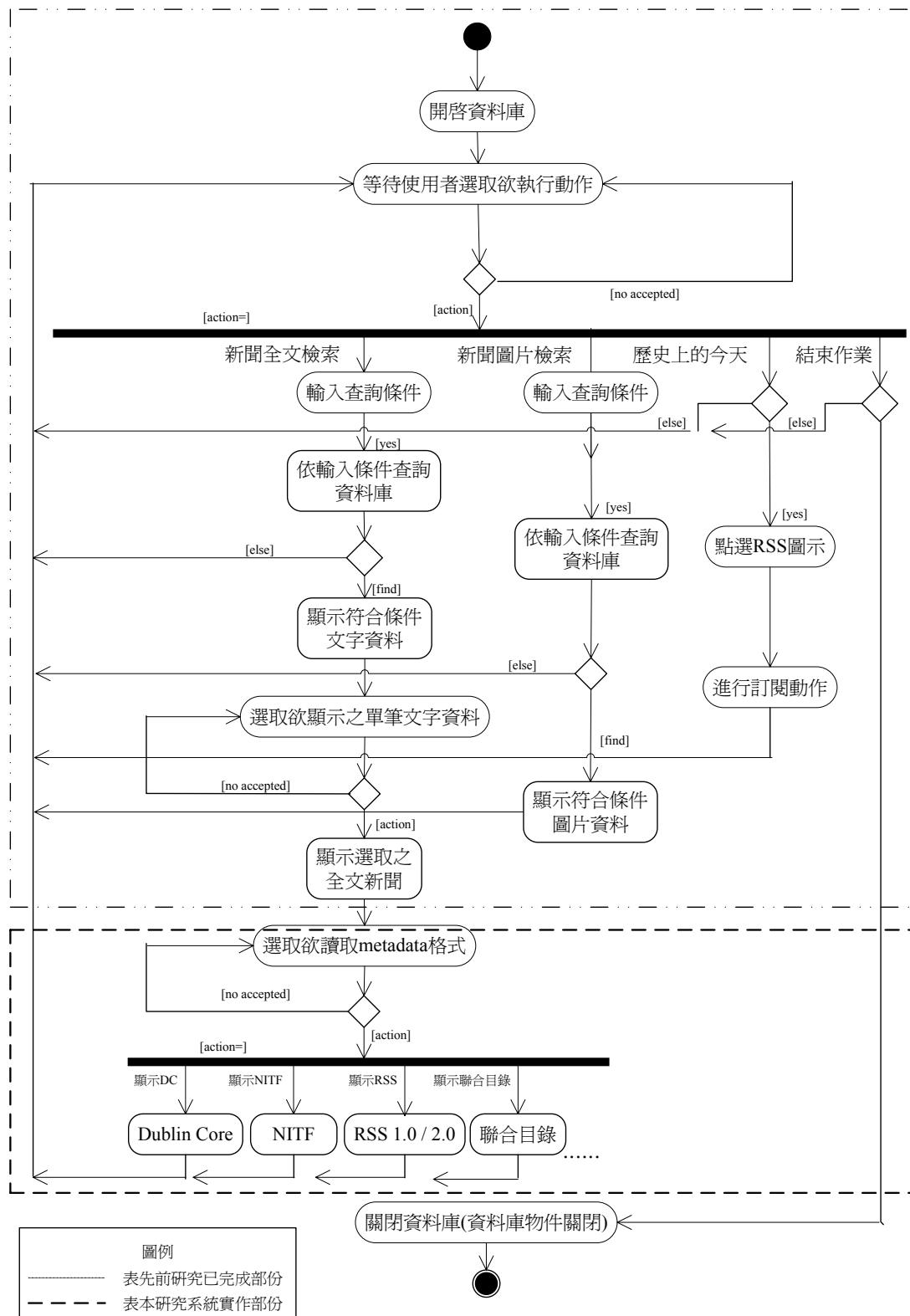


圖 12 檢索子系統之活動圖

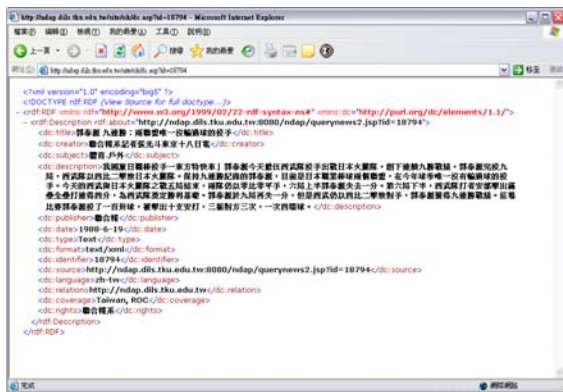


圖 13 RDF+DC 格式轉換結果

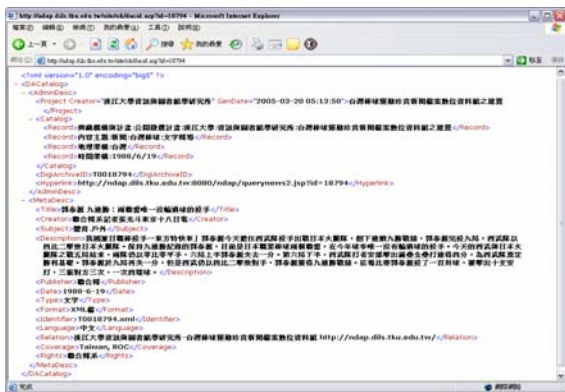


圖 14 DAC 格式轉換結果



圖 15 RSS 1.0 格式轉換結果

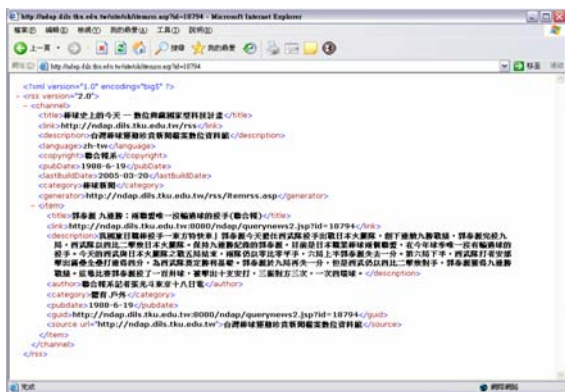


圖 16 RSS 2.0 格式轉換結果

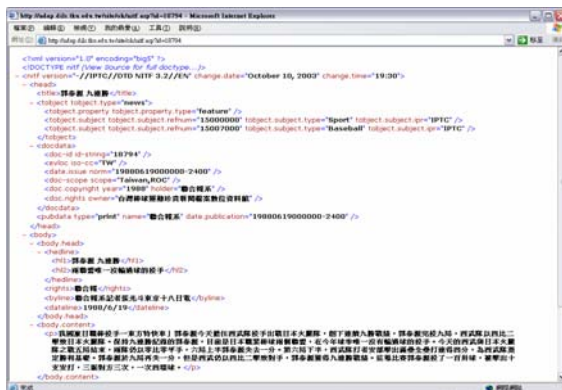


圖 17 NITF 格式轉換結果

4.3 OAI-PMH 檢索系統設計

雖然本系統之大批匯出模組已能順利將Metadata匯出至聯合目錄，然而，聯合目錄之終極目標是所有參與數位典藏的計畫皆能採用OAI-PMH架構，以便完成即時、全自動、高效率的資料彙集。有鑑於此，我們於是接著進行符合OAI-PMH協定之系統開發工作。

而OAI-PMH雖可支援多種Metadata標準，但DC為其之必備格式，稱為OAI_DC，故將參考表 3之新聞Metadata轉換與對應結果加以延伸，將本資料庫內部Metadata欄位與OAI_DC各元素進行對應分析，歸納元素與欄位之登錄內容，以便回應符合OAI-PMH所規範之Metadata記錄集。

4.3.2 實作與驗證

乃基於Metadata轉換系統與OAI-PMH 2.0版本之規範，本計畫開始著手建置臺灣棒球新聞檔案數位資料館之Data Provider伺服器，以支援數位典藏聯合目錄依Service Provider規範所傳送之命令集及參數。

並應用前述對於資料庫之新聞事件設計之棒球新聞類目，來滿足OAI-PMH中之「資料集」元件，應用於聯合目錄之「分類架構」。系統主要功能是剖析（parse）六個命令集及其參數，再依前述新聞類目作為紀錄集，以OAI-PMH定義之XML Schema格式封裝新聞Metadata作回應；此外，因資料量極大，也一併考慮流量控制（Flow Control）問題；而為了能與聯合目錄的Service Provider順利介接，也進行了系統驗證。

4.3.2.1 命令集剖析與回應

針對六個命令集及其參數的剖析，本文以個別活動圖說明其運作原理，實作的回應結果則可直接參閱本計畫網站之實際輸出。

(1) Identify

Identify乃為傳回該Repository的識別資訊，並無伴隨其他參數，因此剖析過程單純，只有當命令錯誤時，回應badVerb之錯誤訊息，其活動圖可參考圖 18。以本計畫為例，repositoryName注錄「臺灣棒球運動珍貴新聞檔案數位資料館」，而baseURL則為「<http://ndap.dils.tku.edu.tw/site/oai/oai.asp>」，earliestDatestamp則為收錄新聞資料的最早日期「1968-01-01」，採用之協定版本protocolVersion則為2.0版。詳細回應結果可連至本系統取得：<http://ndap.dils.tku.edu.tw/oai/oai.asp?verb=Identify>。

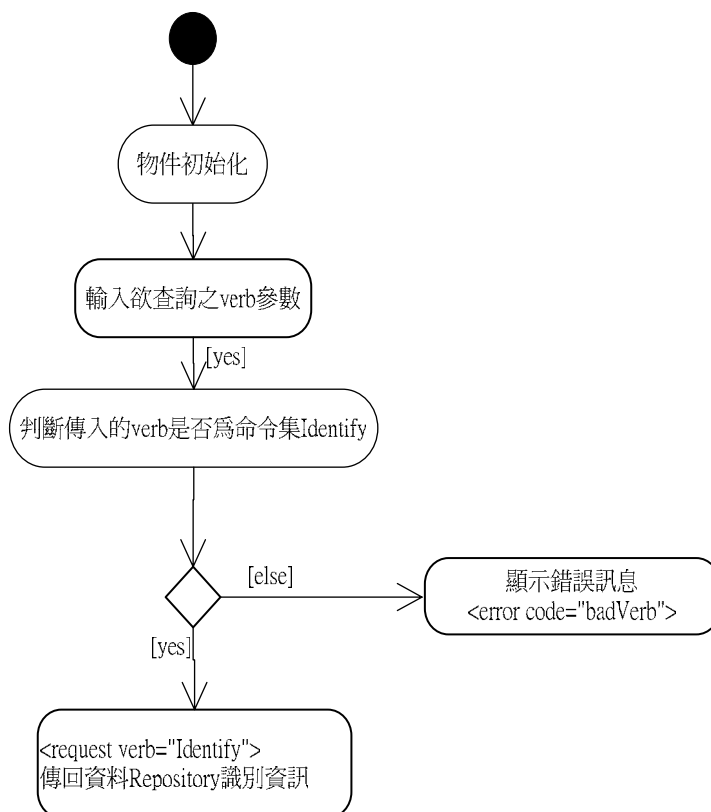


圖 18 Identify 活動圖 (資料來源：本計畫繪製)

(2) ListMetadataFormats

ListMetadataFormats是用以查詢該Repository所支援之Metadata格式，若無指定特定之紀錄，則顯示所有支援之Metadata格式；若有指定單筆紀錄，則顯示該筆紀錄所支援之Metadata格式，圖 19 為其活動圖，實作結果可連至本系統取得：<http://ndap.dils.tku.edu.tw/oai/oai.asp?verb=ListMetadataFormats>。本計畫目前只支援 oai_dc一種Metadata格式，將來可再予以擴充。

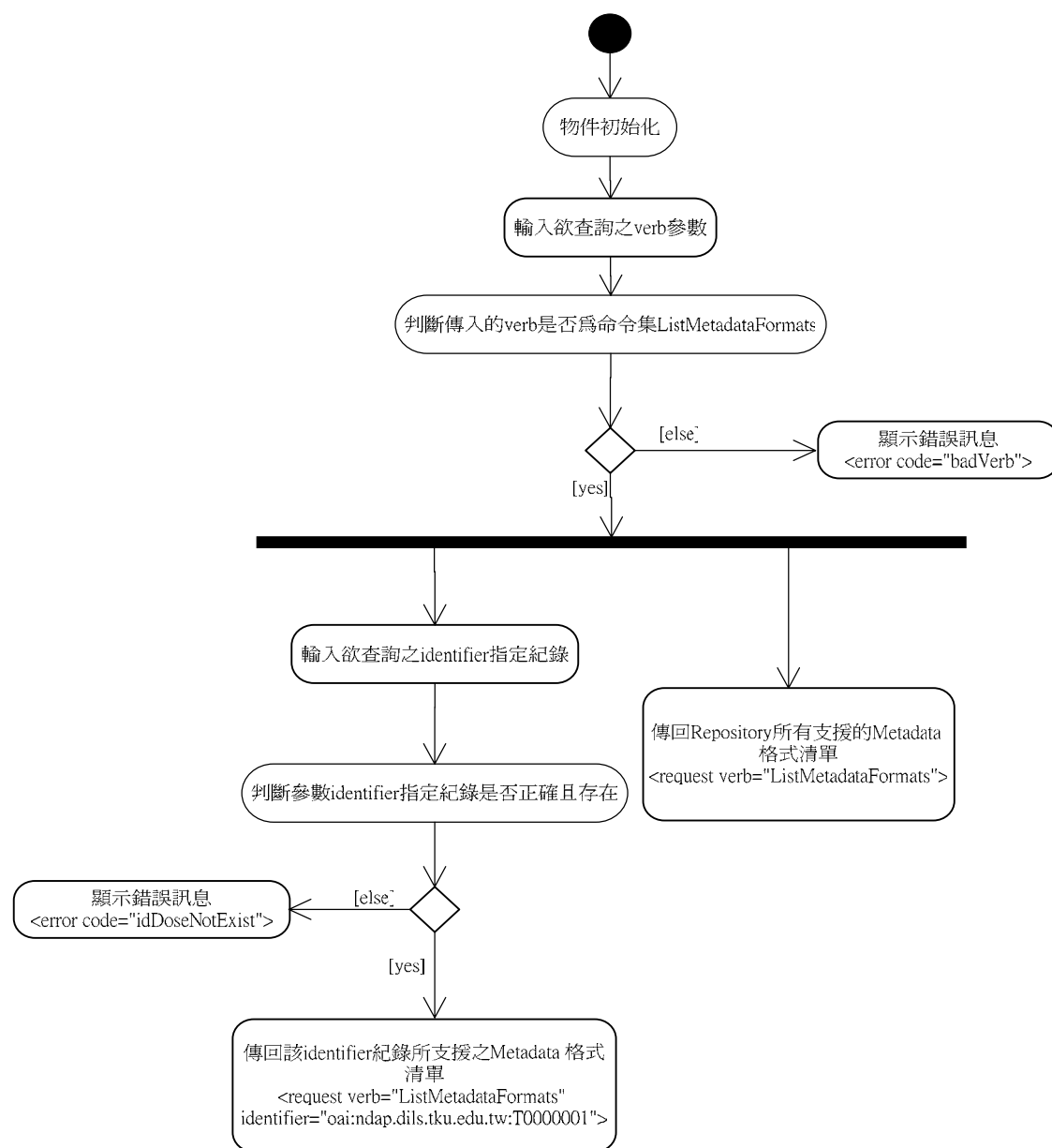


圖 19 ListMetadataFormats 活動圖 (資料來源：本計畫繪製)

(3) ListSets

ListSets為該Repository中所有「資料集」清單，亦即資料儲存器中所有分類類目，setSpec為該類目之標誌，如有複分須遵循OAI-PMH協定以「:」區別下層類目，setName為類目名稱，其活動圖參見圖 20。以本計畫為例，其採用單層類目如圖 2所繪，因未超過15類，故不提供控制流量參數resumptionToken，回應結果可連至本系統取得：

<http://ndap.dils.tku.edu.tw/oai/oai.asp?verb=ListSets>。

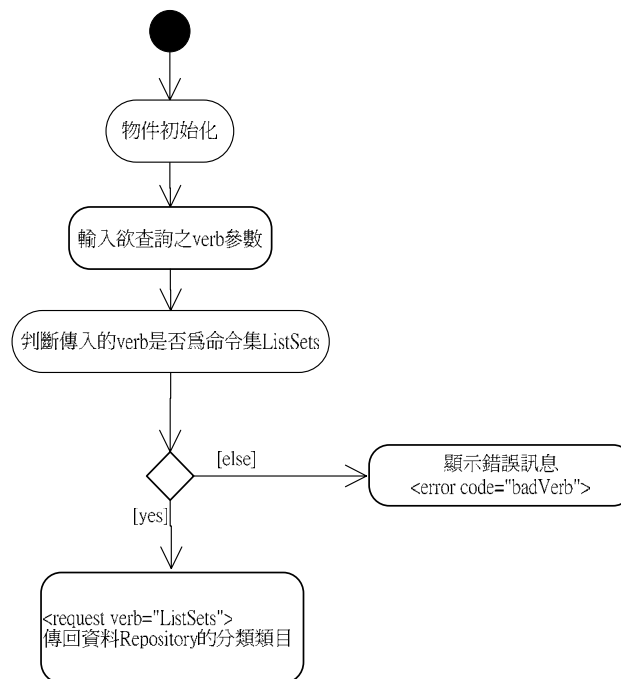


圖 20 ListSets 活動圖 (資料來源：本計畫繪製)

(4) ListIdentifier

ListIdentifier為取得Repository中紀錄標記，其必備參數為metadataPrefix或流量分頁參數resumptionToken，其中有可進階限制回傳資料範圍之參數，若查詢之參數錯誤將回傳錯誤訊息，其剖析流程可參考圖 21。以 http://ndap.dils.tku.edu.tw/oai/oai.asp?verb=ListIdentifiers&metadataPrefix=oai_dc為例，此URL乃為要求回傳metadataPrefix為oai_dc之紀錄Metadata格式，未進階限定範圍，而Repository中之紀錄過多，故顯示流量控制參數resumptionToken，讓使用者可以再次查詢接續之新聞紀錄。

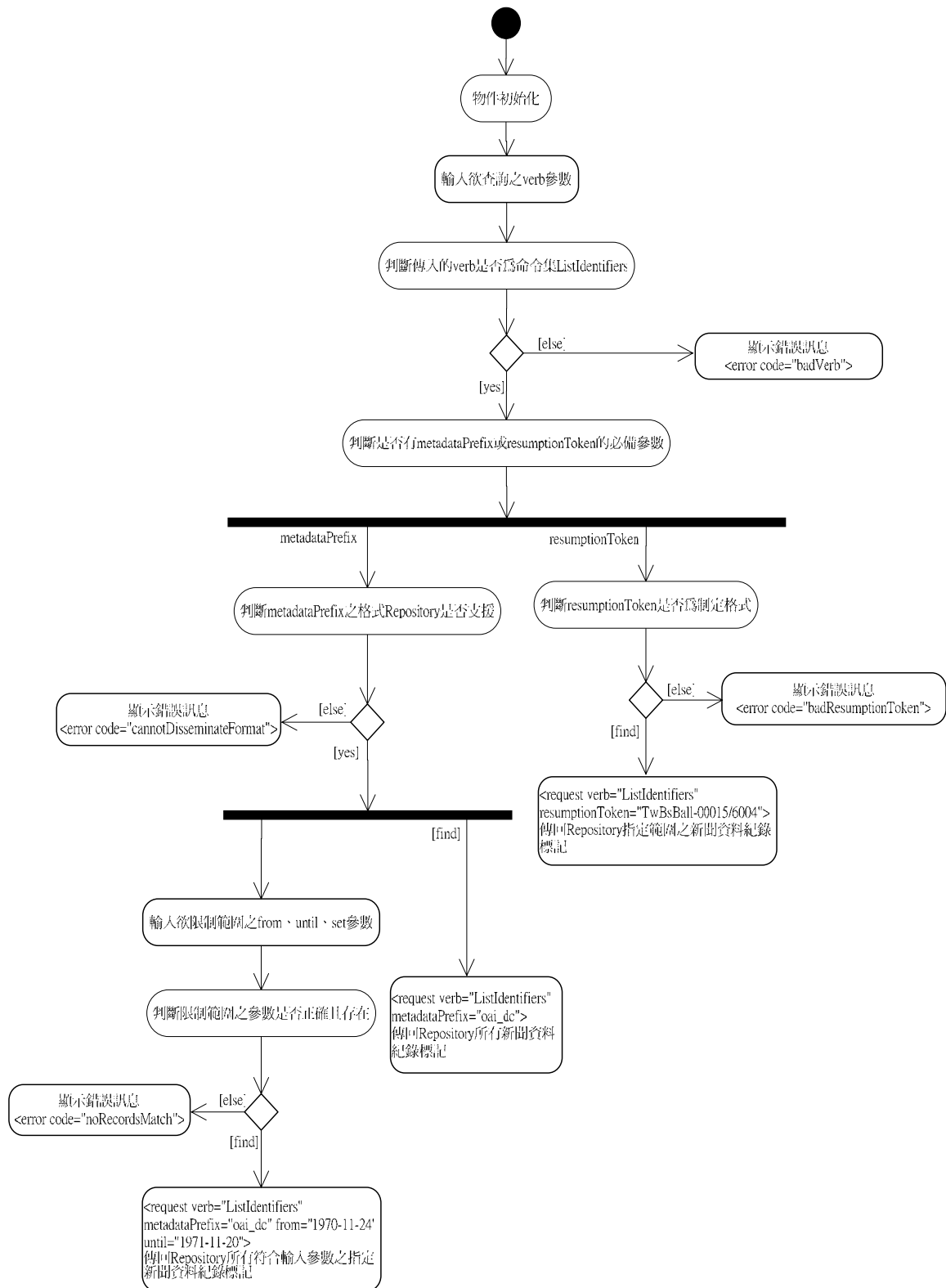


圖 21 ListIdentifiers 活動圖 (資料來源：本計畫繪製)

(5) ListRecords

ListRecords乃用以回傳Repository中指定範圍的所有紀錄明細，其必備參數與可選參數同ListIdentifier之規定，其差異在於回傳之資料為完整明細，而不同於ListIdentifier只有資料之標記部份，其剖析活動圖如圖 9所示。本計畫依表 3中所分析之內部

Metadata與oai_dc對應項目，將新聞事件以OAI-PMH規定之格式回傳，例如：
http://ndap.dils.tku.edu.tw/oai/oai.asp?verb=ListRecords&metadataPrefix=oai_dc，是要求回傳紀錄清單，而無進階限定查詢範圍，因Repository中之紀錄過多，故顯示流量控制參數resumptionToken，讓使用者可以再次查詢接續之新聞紀錄清單。

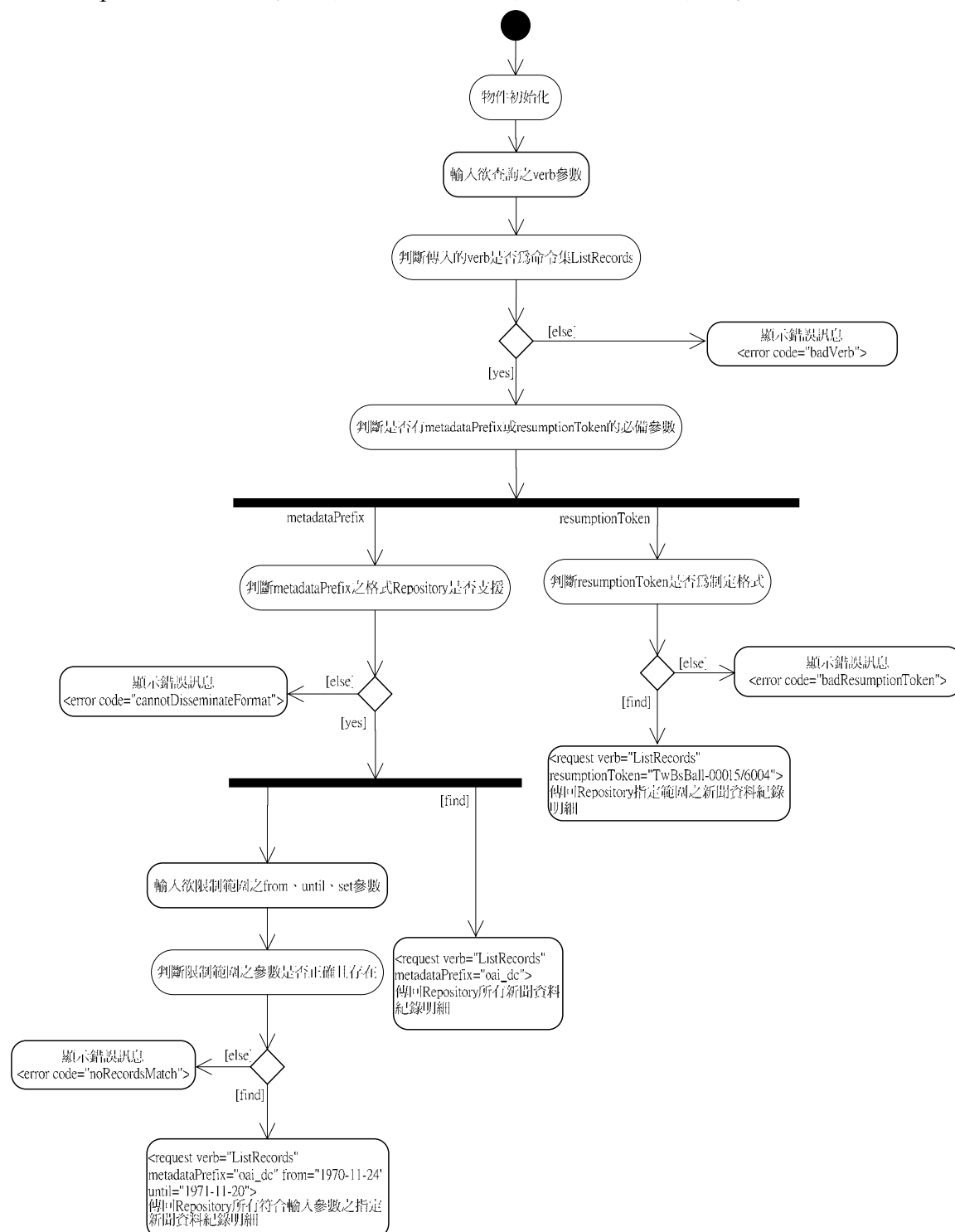


圖 22 ListRecords 活動圖 (資料來源：本計畫繪製)

(6) GetRecord

GetRecord用以檢索Repository中單筆符合查詢範圍之完整Metadata格式紀錄，

Identifier與metadataPrefix為必備參數，以Identifier作為查找Repository之憑據，而依metadataPrefix之指定Metadata格式顯示，其剖析流程可參閱圖 23之活動圖。本計畫之Identifier格式採用「oai:ndap.dils.tku.edu.tw:新聞事件之autoid」，查找符合該Identifier的新聞事件後，依表 3所整理與OAI_DC對應之項目，以XML格式回傳該指定Identifier的紀錄明細。例如 `http://ndap.dils.tku.edu.tw/oai/oai.asp?verb=GetRecord&Identifier=oai:ndap.dils.tku.edu.tw:T0000001&metadataPrefix=oai_dc` 為查找Identifier為oai:ndap.dils.tku.edu.tw:T0000001，指定以OAI_DC的Metadata格式回傳紀錄之結果。

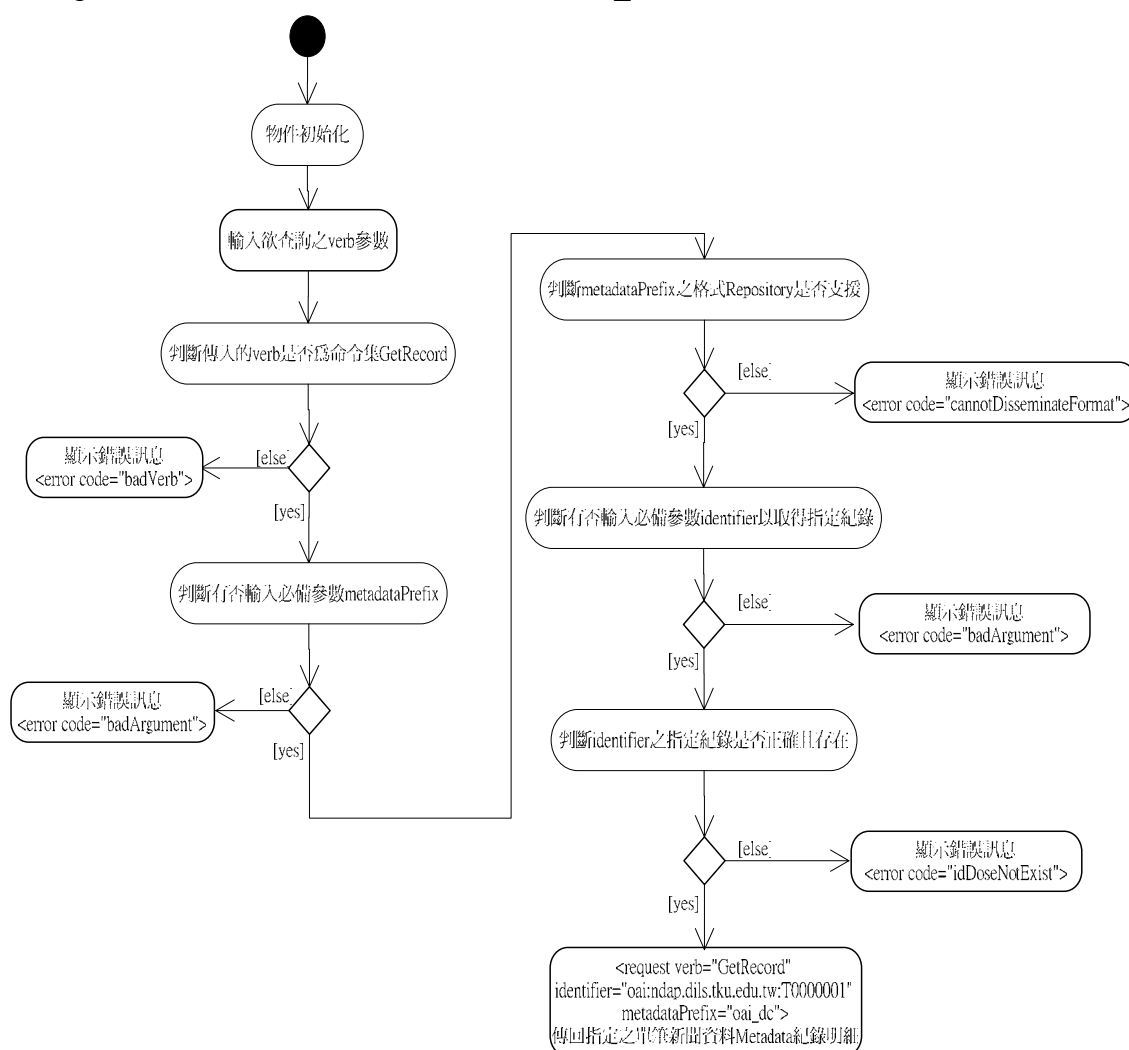


圖 23 GetRecords 活動圖 (資料來源：本計畫繪製)

4.3.2.2 流量控制

當資料庫資料過多時，流量控制可用於分批取得資料，OAI-PMH協定中控制流量之參數為resumptionToken，但並未硬性規定此參數之注錄格式，而由各系統設計者彈性運用。本計畫分析OAI官方網站中已註冊之Data Provider對於此參數之注錄格式，歸納出三種主要方式：

(1) 號碼牌：由Data Provider給予檢索端一個號碼牌作為註記，檢索端可以resumptionToken參數回傳該號碼牌進行再次檢索時，Data Provider則回應檢索端接續於上次檢索的紀錄。號碼牌注錄方式有給予隨機號碼、流水號或檢索時間等方式。

(2) 資料錄編號：Data Provider回應所有符合檢索之資料錄總數，與目前查詢回應之第一筆資料錄為全部資料錄之第幾筆，亦即以「當前資料錄/全部資料錄」之格式為之。

(3) 分頁編號：Data Provider將回應資料錄加以分頁顯示，告知檢索端目前為第幾頁，注錄方式通常為各Data Provider自訂之頁碼格式。

將此三種方法從便利、易懂等角度加以評估後，本計畫決定採取「資料錄編號」之方式，利用「當前資料錄/全部資料錄」之資料錄編號方式管控流量，對於檢索端而言較為容易理解、使用。

4.3.2.3 系統驗證

Data Provider建置完成需通過系統驗證之過程，才能確保符合OAI-PMH協定的規範，也才能與Service Provider順利介接，本計畫利用OAI-PMH官方網站所提供的線上驗證功能³⁵進行檢測。起初一直無法通過OAI驗證系統，雖經研究者不斷修改程式與自行檢驗，確認並無問題卻仍無法順利過關。經由反覆的研究，最後推測可能因本計畫之資料庫內容為中文繁體，採用Big5編碼，而OAI線上驗證系統並未支援此項編碼格式。於是我們另行建置英文版測試資料庫，更改XML編碼為Unicode (UTF-8)，其餘回應之結構、語法與格式則維持不變，再次進行驗證後，便可完全通過其檢測。由此可確定本計畫建置之Data Provider其回應結果符合OAI-PMH協定之規範。

第五章 Wiki 協作系統的應用

5.1 Wiki 協作金三角模式

為了以Wiki協作系統作為知識管理平台，對數位典藏內容進行更進一步的加值利用，我們針對數位化時代之知識型態提出一個適用於Wiki協作環境的金三角模式（Triangle Model），如圖 24所示。

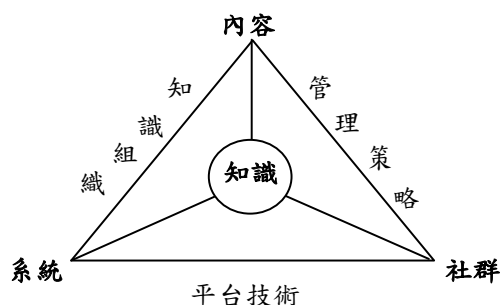


圖 24 Wiki 金三角模式

在此Wiki金三角模式中，知識的創造乃是由「Wiki三要素」——「系統」、「社群」

³⁵ Open Archives Initiative, "Registering as a Data provider OAI-PMH version 2.0", available from <<http://www.openarchives.org/data/registerasprovider.htm>> (2005/07/14)

與「內容」三者，並藉由「Wiki三面向」—「平台技術」、「知識組織」和「管理策略」加以串接，形成一個完整體系共同形塑而成。可表示成如下方程式：

$$\text{Wiki 金三角} = \text{系統} \cdot \text{平台技術} + \text{內容} \cdot \text{知識組織} + \text{社群} \cdot \text{管理策略}$$

實際上，此模型不僅適用於 Wiki 協作環境，對任何虛擬社群的運作皆適用。

以下我們深入探討在建構Wiki協作環境時，如何平衡Wiki金三角中的各個要素與面向，使其充分運作，達到智慧集結與知識分享的目標，並以我們於今年度計畫中所建構的「台灣棒球維基館」（網址：<http://twbsball.dils.tku.edu.tw>，以下簡稱「本站」）作為案例來說明。

5.2 系統建置與平台技術

「台灣棒球維基館」旨在以本計畫之數位新聞資料為基礎，延伸建構一個線上開放的台灣棒球歷史教育館，以記錄台灣棒球發展史中值得記錄的人物、事件、時代、地點、文物 ... 等，並以 Wiki 開放協作方式供大眾對新聞報導之外的相關資料加以彙整，以充實我國棒球發展的歷史紀錄。為達此一目標，研究小組首先進行Wiki協作系統平台建置作業，著手研究Wiki三要素中的「系統」與三面向中的「平台技術」。

為了使系統能更有效地運作，在系統設計、開發、運作等各階段，評估均是不可或缺的工作。而本計畫是採以Wiki協作系統作為平台，是經由系統開發時所需花費的時間、成本、人力及技術等面向作為考量，並在分析幾種可行方案並加以評估後之結果(如表 4)。

表 4 系統開發之評估要素

方式 評估 要素	單純網頁製作	自行開發網頁、資料庫、程式設計	採用 Open Source (如: Wiki、Blog)	購置現成 知識管理系統
時間	高	高	低	中
成本	低	高	低	高
人力	中	高	低	低
技術	中	高	低	低

由上表可得知，開放原始碼(Open Source)是系統開發時眾多方案中較合宜的選擇，然而近年來Open Source型態的社會性軟體（Social Software）發展日益膨脹，使得網路的應用層面越來越人性化且大眾化，其中Wiki與Blog帶領出另一種知識的組織與呈現方式，而Blog比較著重於個人網誌的功能，比較類似於討論版的運作模式，因此較不適合本計畫的運用。而Wiki系統除了具備成本最低、不需開發、簡易架設等節省時間與人力的上述優點外，再加上該系統具備較完善的多元分類、組織與易於管理的功能特性，

並結合超連結、重定向、可搜尋、界面親和度高等特點，最重要的是可透過協同創作的過程來彌補研究小組棒球知識之不足，因此我們選用此套軟體作為技術平台，來作為匯整台灣棒球新聞與其相關資料之加值應用的研究。

因此，運用Wiki協作系統協同創作的特性與知識組織技術的相互結合，為數位內容進行分類、管理與加值應用，如分類(Classification)、索引(Index)、索引典(Thesauri)與控制詞彙(Controlled Vocabulary)等概念，建構一個主題式社群：「台灣棒球維基館」³⁶(如下圖 25)，目前運作已略有規模。



圖 25 台灣棒球維基館首頁

經由反覆評估目前常見的數種 Wiki 引擎，包括 MediaWiki、TikiWiki、Oddmuse、PmWiki、MoinMoin、PhpWiki、TWiki、UseMod 等，從功能完整性、使用難易度、中文支援力 … 等因素加以比較分析，所得評估結果如表 5 所示。最後我們選用了 MediaWiki 作為系統平台，它因被用來撰寫號稱全球最大的線上百科工程「維基百科」(Wikipedia)³⁷而頗具盛名，該系統功能特性亦符合本計畫之應用。目前本計畫採用的規格如下：

- 軟體版本：MediaWiki 1.44；
- 程式語言：PHP；
- 資料庫管理系統：MySql + MysqlAdmin；

³⁶ 淡江大學資圖所數位典藏計畫小組，台灣棒球維基館，<<http://twbsball.dils.tku.edu.tw>>。

³⁷ Wikipedia, "Wikipedia", available from <<http://wikipedia.org/>>(2006/3/5)

- Web 伺服器：Apache。

表 5 Wiki 系統評估總表

系統功能 維基引擎	程式語言	資料存取方式	系統功能強弱	系統建置難易	系統普及率	系統更新頻率	中文支援能力分析					中文能力綜合評估	推薦等級
							預設文字編碼	預設操作界面	支援中文操作介面	支援中文內容顯示	支援中文頁面連結		
MediaWiki	PHP	MySQL	強	易	高	高	UTF-8	英文	有 (安裝時選擇)	有 (預設即支援)	有 (預設即支援)	高	5 ⁺
TikiWiki	PHP	MySQL	強+	中	中	中	UTF-8	英文	有 (安裝後選擇)	有 (需修改語系設定即支援)	有 (預設即支援)	高	5
Oddmuse	Perl	TEXT	中	易	中	低	UTF-8	英文	有 (需擴充程式)	有 (需由擴充程式)	有 (預設即支援)	中	4
PmWiki	PHP	TEXT	中	中	中	中	ISO-8859-1	英文	無	有 (需修改語言編碼設定)	有 (需擴充程式)	中	4 ⁻
MoinMoin	Python	TEXT	中	中	高	中	UTF-8	中文	有 (預設中文介面)	有 (預設即支援)	有 (預設即支援)	高	3
PhpWiki	PHP	MySQL	中	中	中	低	ISO-8859-1	英文	有 (需擴充程式)	有 (需修改語言編碼設定)	有 (需修改語言編碼設定)	中	3
TWiki	Perl	TEXT	強	難	高	中	ISO-8858-1	英文	無	有 (需修改語言編碼設定)	有 (需修改語言編碼設定)	中	2
UseMod	Perl	TEXT	弱	易	高	低	ISO-8859-1	英文	有 (需擴充程式)	有 (需由擴充程式)	無	低	1

5.3 知識組織與內容加值

5.3.1 分類架構的設計

分類是依據物件間的關係將其排序分組的一種活動³⁸，John Dewey 更曾提出分類就是知識（all knowledge is classification）的理念³⁹。因此，也說明了分類架構在知識體系的範疇中是極具關鍵性影響力的。

本計畫針對棒球運動相關主題，進行樹狀分類架構的設計，如圖 26 所示，並利用 MediaWiki 系統中 category 的功能將同類之內容組織起來，即可將相同類別的頁面匯集成

³⁸ Sherry Koshman. "Categorization and classification revisited: a review of concept in library science and cognitive psychology." Current Studies in Librarianship Spring/Fall 1993: 26.

³⁹ Bo Leuf, Ward Cunningham, The Wiki Way: Quick Collaboration on the Web, Addison Wesley, Addison Wesley, 2001.

一個類別集合。此外，本系統也具備可彈性製作多元分類的能力，每個頁面可依不同屬性歸類於不同類別中。以棒球人物「郭泰源」為例，可輕易將其歸類為人物、球員、教練、旅日球員、成棒時期人物、職棒時期人物、棒壇重要人物等多元分類。

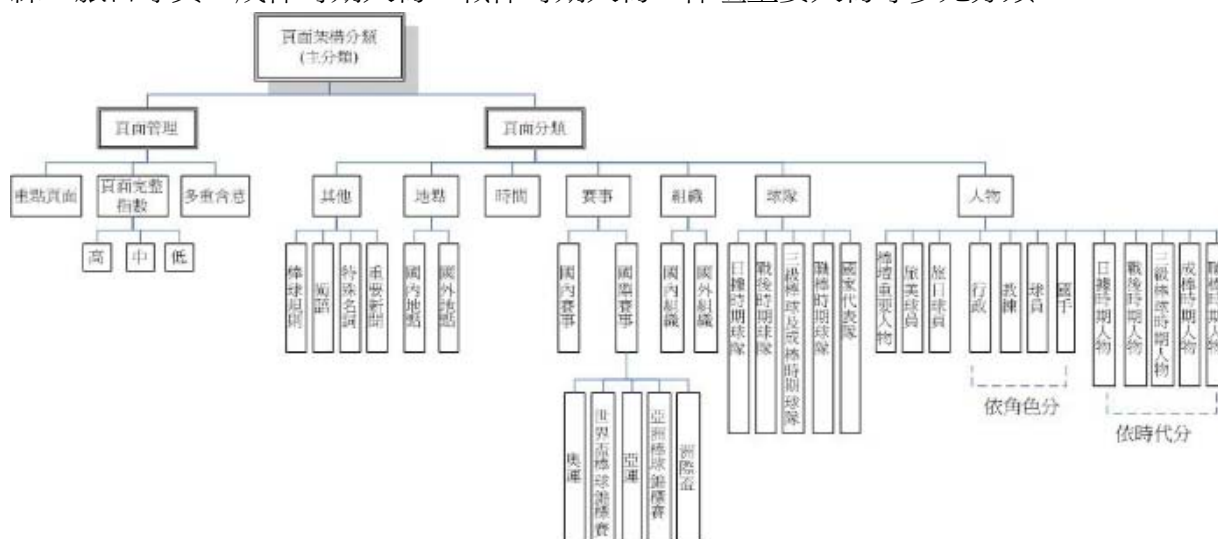


圖 26 本站內容分類架構圖

5.3.2 索引概念的應用

索引係將資料中所提及的各詞彙，包含人、事、時、地、物、主題或概念等提出，再依照特定的排列順序（如筆劃、字順等）列出，並註明資料來源以作為指引，以便取得原始的資料或全文。我們可透過系統中的二項功能達到此目的：其一是「列出所有頁面」的功能，系統會按照內碼順序列出所有頁面；其二是可藉由分類方式，依筆劃順序條列該類別所屬頁面。兩者皆是透過索引連結，指引使用者找到資料來源。

5.3.3 控制詞彙的製作

根據美國國家標準ANSI/NISO Z39.19-1993 (Guidelines for the Construction, Format, and Management of Monolingual Thesauri)⁴⁰指出，控制詞彙可清楚地展現詞彙間同義、同形異義、層級和關聯的關係，其目標是促進文件標引的一致、控制後組合資訊儲存與檢索系統、藉由款目詞與描述語的連結而便於搜尋⁴¹。本計畫作法是借用類似主題地圖 (Topic Maps) 之TAO概念⁴²，將與棒球相關之各主題 (Topics，如球隊、人物、組織等) 先行萃取出來，定義其相互關聯性 (Associations)，並提供概念之間的關係，讓彼此間的相互關係更為清晰以發揮更大效用，並指引至目標處加以呈現 (Occurrence)，也可同時揭示該特定領域的標準用語，透過Wiki協作系統的運作，再搭配詞彙分類組織、同義詞、階層與關聯概念的製作，使其在知識組織概念上的運用更臻完善。

- 同義的概念：運用系統中「重定向」(redirect，又稱「轉向」) 功能 (如下圖 27)，

⁴⁰ ANSI, "ANSI Z39.19-1993(R1998) Guidelines for the Construction, Format, and Management of Monolingual Thesauri", available from <<http://www.niso.org/standards/resources/Z39-19.pdf>>(2005/03/22)

⁴¹ 同註 25。

⁴² TopicMap.Org, "XML Topic Maps (XTM) 1.0 - TopicMaps.Org Specification," 8 Aug. 2001, <<http://www.topicmaps.org/xtm/1.0/index.html>>

將相同意義的詞彙，指向共同的目標，建立其參照關係。此種方式可將概念上或用法上相同，或視為相同的詞彙串聯起來，有效達到控制詞彙統一術語的一致性目的，在檢索時就特別能發微其效用，例如：「東方特快車」轉向「郭泰源」、「亞洲巨砲」轉向「呂明賜」、「微笑喬治」轉向「趙士強」、「黃金戰士」轉向「陳致遠」、「鐵血教頭」轉向「曾紀恩」…等，皆為同義概念的應用。

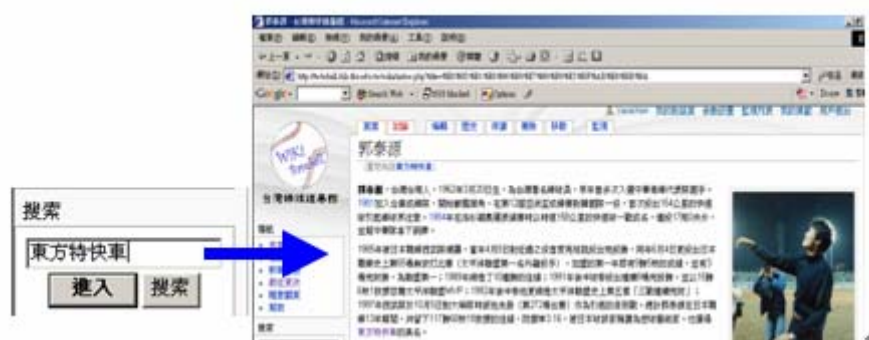


圖 27 控制詞彙的製作 — 同義的概念

- 階層與關聯的概念：欲實現階層與關聯的概念，本計畫採用的方式是以主題詞為標題，在其Wiki頁面中，將相關詞彙列出，再描述其廣義詞、狹義詞或相關詞，以呈現其上下層主題間的關係，並透過連結的功能將彼此串聯起來。如此可提醒使用者有另一種關係的存在，並引導使用者加以利用。例如：「紅葉」為「紅葉國小」、「紅葉少棒隊」的廣義詞；「紅葉國小」、「紅葉少棒隊」則為「紅葉」的狹義詞；「紅葉國小」、「紅葉少棒隊」兩者則互為相關詞。
- 同形異義的概念：在知識組織過程中，時常遇到同形異義的情形，本計畫的作法是採用另加一個「多重含意頁」的方式處理（如下圖 28）。例如：台灣棒壇人物中曾有兩位「林建良」，同名同姓卻不同人，因此先以「多重含意頁」說明以便判讀，再採間接連結方式將其區隔取用，搜尋時也能有所區別。



圖 28 控制詞彙的製作 — 同形異義的概念

5.3.4.綜合應用

除以知識組織技術的實踐外，在「台灣棒球維基館」中更進一步將內容加值成果予以綜合應用，例如：

- 台灣棒球發展史：台灣棒球發展史是以電子書方式訴說台灣棒球運動發展的軌跡，讓讀者可像閱讀書籍般，並在資訊架構中搭配上下單元的情境式導覽設計，導引讀者依章節閱讀相關資訊。
- 台灣棒球大事記：以年度為單位，依據年份排列整理台灣棒壇大事記，並於文中運用變數引入功能輔以超連結將相關資訊作更詳盡的說明及建構。
- 台灣棒球名人錄：將台灣棒球史上各個不同時期的重要人物加以彙整，按照年代加以排列，並附上簡要說明，以增加頁面的可讀性。
- 台灣棒球影像館：此區域展示具歷史價值的照片，包括人物、球隊、場地、賽事、球具、球裝、球卡...等，將其以相簿的形式呈現。
- 台灣棒球重要新聞報導：收錄由聯合報授權提供的歷史性棒球新聞，並呈現其新聞標題、記者、日期、來源、版數、內文等，運用超連結功能可將新聞中所提及相關的人、事、時、地、物的概念串連起來，讓知識的呈現形成網絡的形態。

5.4 社群建構與管理策略

當「台灣棒球維基館」的「系統」與「內容」雛形完成後，接著我們便開始思考Wiki金三角中的第三要素：「社群」和第三面向：「管理策略」的運作。然而在網路的環境中，如何經營一個具歸屬感的虛擬社群，使數位內容維持穩定的成長，因此在虛擬社群的經營、社群管理策略、社群規範、頁面品質管理、數位內容版權管理、管理者引導技巧、社群激勵機制及社群推廣與行銷等面向的探討，都會影響其數位內容之品質及使用率。因此應該依據實際狀況而加以修正，這也是維持社群正常運作的重要環節之一，如此一來，在吸引社群成員之餘，才能增加其參與度，並建立其對社群的信賴及忠貞度，以達到社群的滿意度。

5.4.1 虛擬社群經營

Armstrong & John Hagel 指出虛擬社群是在網路的環境中，經由社群成員針對一個特定的主題持續經營，以使社群得以不斷地成長擴大。網路虛擬社群通常是經由線上群組成員分享某行為或某主題，試圖找到一個群體，使群體內部的成員能夠共存，以及使彼此之間更加的親近。⁴³Rheingold認為虛擬社群係源自於電腦中介傳播所建構而成的虛擬空間(cyberspace)，是一種社會集合體(social aggregation)，它是發生來自於虛擬空間上有足夠的人、足夠的情感、與人際關係在網路上長期發展。⁴⁴虛擬社群不僅能讓社群成員透過網際網路互相溝通與資訊分享，更重要的是它把人們聚集在一起，共同分享資訊，甚至產生並吸收社群成員所創造的知識。持續地互動，並從互動中創造出一種相互

⁴³ Armstrong & John Hagel (1997), "Net Gain: Expanding Markets Through Virtual Communities."

⁴⁴ Rheingold, H.(1993)., "The virtual community .", MA:AddisonWesley.

信賴(mutual trust)與相互仰賴(mutual dependency)的夥伴關係(partnership relationship)。⁴⁵

Chris 等人提出虛擬社群六大構成要素，如圖 29所示，虛擬社群應以成員之間的歸屬感為中心，由以下六個要素來強化：提供珍貴豐富的内容(Precisely tailored content)、社群的認同(Identification with the brand)、對其他成員有強烈志同道合的感覺(Awareness of other Like-minded users)、彼此的互動(Ability to interact with others on website)、參與發展的機會(Opportunity to shape the development of website)、參與虛擬社群所產生或擁有共同利益(Mutual Benefits of participation)。⁴⁶

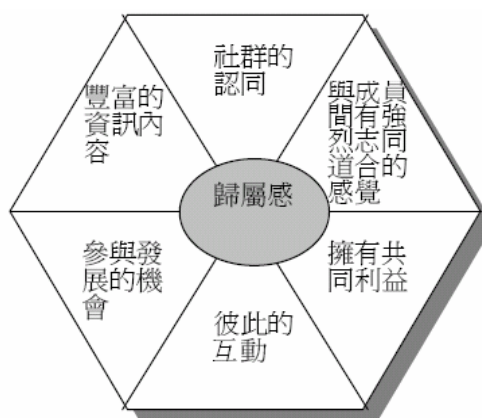


圖 29 虛擬社群的六大構成要素

因此，一個令社群成員滿意度高的虛擬社群也可說是一個匯集知識、建立成員夥伴關係的虛擬社會。由上述理論之說明可得知，台灣棒球維基館是個以台灣棒球相關議題作為主題，並集合熱愛棒球的社群加以共同協作，是虛擬社群的實際運作模式之一。

5.4.2 Wiki 社群結構

在Wiki協作環境中，社群主要是由讀者、作者與管理者所組成。讀者是指僅閱讀資料而無進行任何編輯動作的使用者；作者則不僅止於閱讀，而是實際參與協作編輯的核心使用者，更是Wiki數位內容之重要來源，任何讀者只要願意編寫，隨時皆可成為作者，不受任何條件限制；管理者則是具有文章刪除/恢復、帳號封鎖/開放、系統設定/維修等能力的超級使用者，具有較高的權限，以維護系統正常運作。久而久之，一群同好便會形成一個網路討論群，形成一種基於資訊分享與情感支持的專屬網路虛擬社群文化。

5.4.3 社群管理策略

社群管理者在社群的互動與發展有著密不可分的關係，更是扮演核心角色。它是指建構與經營虛擬社群的人，通常需同時扮演多種角色，負責包含如系統、內容、討論經營等工作項目。Pór 也指出虛擬社群經營者在社群扮演系統管理員（system administrator）、權限管理員（account administrator）、社群建築師（community architect）、

⁴⁵ Romm, C., Pliskin, N. and Clarke, R., 1997, "Virtual Communities and Society: Toward an Integrative three phase model", International Journal of Information Management, Vol. 17, No. 4, p. 261-270.

⁴⁶ Chris, M., Miriam, M., Kevin, O. and Ashok, G., 1999, "Making Real of Virtual Community", Pricewaterhouse Coopers

討論主持人 (moderator)、內容促進者 (Content Facilitator) 及網站導航員 (Cybrarian) 等六種角色，這些角色與工作任務在小規模的社群中可以是一個人或幾個人所擔任，在大規模的社群中宜各由專人負責。⁴⁷

5.4.3.1 建立社群規範

社群規範是讓所有成員可遵循的相關原則，目前，「台灣棒球維基館」的管理者制訂了下列幾項規範：

- 一般規範：旨在宣揚本站理念，包括成立宗旨、關於「台灣棒球維基館」、發展歷程、網站日誌 … 等。
- 輔助說明：內含本站編輯原則與手冊，以供協同作者遵循，目前包含的幾項說明，主要針對編輯與命名規則、編輯語法、分類原則、編輯樣版、模板使用、模範頁面及工具列說明 … 等加以說明。
- 免責聲明：因 Wiki 開放協作過程具有「先出版後編輯」之特性，頁面可能出現過渡、不完整或甚至錯誤之內容，因此，我們以「免責聲明」方式進行告知，提醒使用者使用本實驗網站內之資源時必須有此方面之認知。
- 數位版權協議：採用 GNU 自由文檔許可證 (GNU Free Documentation License)⁴⁸ 之方式，以避免版權糾紛。此外，本站中的「新聞資料」與「照片資料」的著作權利仍屬原權利人所有，僅供瀏覽之用，如有他用需取得原權利人授權。

在這些基礎工作完成後，則開始將「管理策略」朝頁面品質管理、社群規模擴增、網站推廣 … 等重點方向進行。

5.4.3.2 頁面品質管理

為了提升頁面品質，我們開始著手於頁面中標示「頁面品質指標」，記載該頁面之完整度與重要指數，或是否有多重含意及頁面需合併的情況；並推動「頁面品質提升計畫」，著手匯集重要性較高，或點閱、查詢頻率較高但內容不完整的頁面，請社群成員共同參與協作編修，同時也針對特定主題推出同類型的一系列頁面，作為週期性改進計畫。

5.4.3.3 管理者引導

- 協助新成員：管理者或資深使用者以模版留言方式，指引新加入成員以熟悉社群的運作模式。
- 社群討論議題設定：管理者妥善運用社群討論功能，尋找能引起使用者興趣的議題，並適當的規劃其結構及內容，以增加社群的互動性與凝聚力。
- 帶動社群討論風氣：管理者給予參與社群討論的成員相關回應，並持續進行討論，達成認同與共識，同時也能進一步瞭解社群成員的反應與意見。

5.4.3.4 社群激勵機制

Hendriks根據Herzberg提出的雙因子理論來探討知識分享動機的相關因素，研究發現人們分享知識的原因較傾向於激勵因子，而非保健因子。以知識分享兩主體分開來看，知識擁有者分享知識的目的是期望能夠獲得他人對工作的肯定與欣賞、可能藉此獲得升遷的機會，或因責任感的驅使而將知識分享給他人，他們也同樣期望分享的行為日後將

⁴⁷ Por, George.(1997)., "The CoIL FAQ on Roles & Responsibilities in Facilitating Online Forums.", Available From< <http://www.co-i-l.com/coil/knowledge-garden/vcfaq/facfaq1.shtml>>(2006/03/16)

⁴⁸ Free Software Foundation, "GNU Free Documentation License", available from <<http://www.gnu.org/copyleft/fdl.html>> (2006/3/5)

獲得互惠的回饋，並創新知識與資訊；而知識需求者可能為處理工作的挑戰、提高作業自主性，以及滿足個人的成就感等因素，而努力吸取知識。⁴⁹有鑑於激勵機制對於社群運作的重要性，並鼓勵參與者的熱心參與，本站制定了「台灣棒球維基館獎勵方式」，依照參與的時間、編輯的次數、新頁面的數目多寡來頒發，只要符合條件即可擁有，並於用戶頁面及「台灣棒球維基館榮譽記錄」中頒予公布名單。目前設制了下列幾項榮譽勳章：

- 新人榮譽：熱心編輯、優秀新人
- 進階榮譽：助理編輯、執行編輯、資深編輯
- 專業榮譽：棒球達人、攝影達人
- 管理榮譽：系統管理員、頁面管理員、見習管理員

此外，另以「本站管理員」、「本站貢獻者」等頁面對努力付出者致謝。

5.4.3.5 社群推廣與行銷

Armstrong & John Hagel曾提出，發展社群會員共可分為發展吸引會員有如下所示的四個階段：吸引會員、增加參與、建立忠貞、及獲取價值等四個階段。⁵⁰由此也可得知，社群發展的第一步便是其行銷策略的規劃，如何宣傳運作才能吸引潛在的社群成員並提高社群能見度，確實是在社群建置後重要的環節。

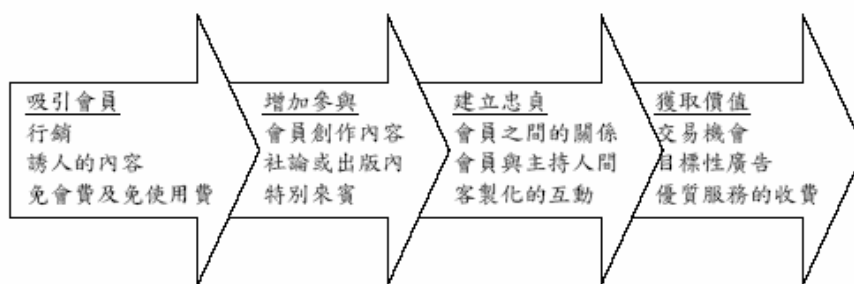


圖 30 吸引社群成員四個階段

然而如何才能提升本站的利用率，目前提出的相關因應措施如下：

- 至各棒球相關論壇及球隊官方網站進行交流。
- 利用電子郵件與電子報傳播。
- 定期參與相關活動，如：研討會、聚會等。

第六章 結論

茲將本計畫今年度的研究成果歸納如下：

(A) Metadata 資訊組織：本計畫網站網址為 <http://ndap.dils.tku.edu.tw>。目前已針對 DC、NRMF、DAC（聯合目錄）、NITF、RSS 等 Metadata 格式進行對應與轉換。

(B) OAI-PMH Data Provider：建置符合 OAI-PMH 協定之檢索系統，以滿足聯合目錄扮演整合式入口的 Service Provider 角色，各個計畫則為分散式 Data Provider，計畫之資料庫則為 Repository 提供 Metadata 檢索服務，整體架構如圖 31 所示。目前已將

⁴⁹ Hendriks, P. (1999)., "Why Share Knowledge? " The Influence of ICT on Motivation of Knowledge Sharing. Knowledge and Process Management, 6(2), p.91-100.

⁵⁰ Hagel III, J. and Armstrong, A. G., 1996, "Net Gain:Expanding Markets Through Virtual Communities", Mckinesy and Company.

第一年、第二年計畫數位化產出結果共三萬五千餘篇棒球新聞以 OAI-PMH 協定與數位典藏聯合目錄系統相互連結傳遞。

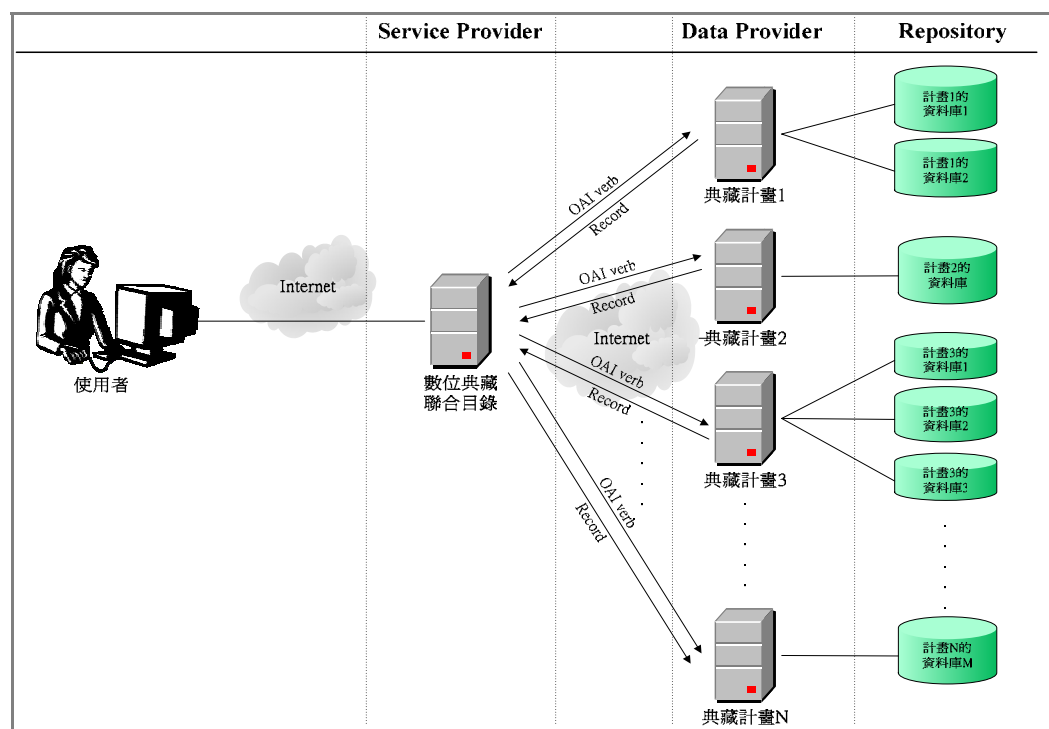


圖 31 數位典藏聯合目錄 OAI-PMH 架構 (資料來源：本計畫整理繪製)

(C) Wiki 加值系統建置：依人、事、時、地、物進行分類架構設計、索引概念應用、控制詞彙製作，並製作出台灣棒球發展史、台灣棒球大事記、台灣棒球名人錄、台灣棒球一級棒、台灣棒球影像館等數位化專輯。

(D) 論文發表：本計畫第二年度至 2006 年 2 月截止共發表七篇學術論文，分別是：

- (1). 林信成、游忠諺、康珮熏，「新聞數位典藏之 Metadata 探討」，21 世紀數位圖書館發展趨勢，頁 23-43，民 94 年 2 月。
- (2). 林信成、康珮熏，「報紙新聞數位典藏 Metadata 轉換系統之設計與應用」，中文媒體數位典藏與新聞標示語言研討會，頁 B2-1~B2-23，台北・國家圖書館，2005/5/11~2005/5/12。
- (3). 林信成、康珮熏，「建置以 OAI-PMH 為基礎的數位化新聞 Metadata 分散檢索系統」，2005 台灣網際網路研討會 (TANET 2005)，台中・中興大學，2005/10/26 ~ 2005/10/28。
- (4). 林信成、陳瑩潔，「Wiki 協作系統在數位典藏內容加值之應用研究」，2005 台灣網際網路研討會 (TANET 2005)，台中・中興大學，2005/10/26 ~ 2005/10/28。
- (5). 林信成、康珮熏，「Metadata 與 OAI-PMH 在新聞數位典藏之整合應用」，教育資料與圖書館學，2005 年 9 月，43(1)，頁 49-68。
- (6). 林信成、康珮熏，「新聞數位典藏之資源共享與開放檢索系統設計」，第三屆演化式計算應用研討會，台北・真理大學，2005/12/03。

- (7). 林信成、陳瑩潔，「從 Wiki 金三角模式探討台灣棒球維基館之知識匯集與分享」，第三屆演化式計算應用研討會，台北・真理大學，2005/12/03。

(E) 會議參與：本計畫定期參與內容發展公開徵選計畫辦公室所召開的工作進度執行報告會議，新聞主題小組所召開的小組討論會議，以及由本計畫團隊自行召開的小組研習會。